

# OPTIMIZATION METHODS ON RIEMANNIAN MANIFOLDS VIA EXTREMUM SEEKING ALGORITHMS \*

FARZIN TARINGOO<sup>†</sup>, PETER M. DOWER<sup>†</sup>, DRAGAN NEŠIĆ<sup>†</sup> AND YING TAN <sup>†</sup>

**Abstract.** This paper formulates the problem of Extremum Seeking for optimization of cost functions defined on Riemannian manifolds. We extend the conventional extremum seeking algorithms for optimization problems in Euclidean spaces to optimization of cost functions defined on smooth Riemannian manifolds. This problem falls within the category of online optimization methods. We introduce the notion of geodesic dithers which is a perturbation of the optimizing trajectory in the tangent bundle of the ambient state manifolds and obtain the extremum seeking closed loop as a perturbation of the averaged gradient system. The main results are obtained by applying closeness of solutions and averaging theory on Riemannian manifolds. The main results are further extended for optimization on Lie groups. Numerical examples on Riemannian manifolds (Lie groups)  $SO(3)$  and  $SE(3)$  are also presented at the end of the paper.

**Key words.** Extremum Seeking Control, Riemannian Manifolds.

**AMS subject classifications.** 34A38, 49N25, 34K34, 49K30, 93B27

**1. Introduction.** Optimization on manifolds is an important research area in optimization theory, see [2, 36, 41]. In this class of problems, the underlying optimization space is a manifold and consequently the analysis differs from the standard optimization algorithms in Euclidean spaces. Numerical techniques and methods for optimization on manifolds should guarantee that in each step an optimizer is an element of the search space which is a manifold. Hence, optimization methods on manifolds are closely related to geometry of manifolds, see [2, 12, 25, 41].

As known, smooth manifolds can be embedded in high dimensional Euclidean spaces (Whitney Theorem) [3]. This means that optimization on manifolds can be carried out as constrained optimization problems in Euclidean spaces with sufficiently large dimensions. However, the corresponding embeddings for each particular manifold may not be available and algorithms developed on manifolds may be more efficient in terms of convergence speed and calculation burden [32]. The algorithms investigated in this field range from simple line search methods to more sophisticated algorithms such as trust region methods, see for example [5, 43].

Optimization on manifolds can arise in a wide range of applications where the search space is constrained. Its applications may appear in signal processing [26], robotics [15], and statistics [10]. The main underlying assumption in most of the numerical algorithms presented for optimization on manifolds is that the cost function is available. This makes the implementation of numerical algorithms simple since various numerical methods can be applied to calculate the sensitivity of cost functions to obtain numerical optimization trajectories. However, in many problems, cost functions may not be given in a well defined closed form. This necessitates generating a class of numerical methods which do not explicitly depend on the closed form of cost functions derivatives [35, 37]. This is the main motivation of this paper.

Extremum seeking is a class of on-line or real-time optimization methods for optimization of the steady-state behavior of dynamical systems [21]. This method is applied for optimization of both static functions and dynamical systems where the

---

\*This work was supported by the Australian Research Council Discovery Project DP120101144.

<sup>†</sup>Department of Electrical and Electronic Engineering, The University of Melbourne, {ftaringoo, pdower, dnesic, yint}@unimelb.edu.au.

optimizer does not have a model of the cost/utility function or dynamical models. In other words, either cost functions or dynamical systems are unknown for the optimization procedure and the optimization algorithm should be able to converge to a vicinity of a local optimizer, see [8, 13, 14, 18, 21, 27–29, 38, 40]. In this paper we only consider the extremum seeking algorithms for optimization of static cost functions and consequently we assume that cost functions are not available for the optimization procedure.

Extremum seeking finds its applications in a vast area of dynamical systems including robotics and mechanical systems. As is known mechanical systems are mathematically modeled on manifolds which do not necessarily possess vector space properties, see [1, 6, 7]. Traditionally, extremum seeking systems have been analyzed in the class of unconstrained optimization methods on  $\mathbb{R}^n$  where the vector space properties of Euclidean spaces simplify the analysis.

In a more general framework, the underlying Euclidean spaces can be replaced by Riemannian manifolds. That is to say, we change an unconstrained optimization problem to a constrained one where the constraints are imposed by the ambient manifold spaces. This necessitates a generalization of the extremum seeking framework for optimization on manifolds. To this end, we define a class of online optimization methods where the optimization trajectories lie on manifolds. As an example, the standard gradient descent and Newton methods are extended to their geometric versions by employing the notion of geodesics on Riemannian manifolds, see [25, 36]. In this paper this step is done for extremum seeking algorithms by introducing the so-called *geodesic dithers* which are the geometric versions of dither signals in standard extremum seeking framework, see [21, 38]. By employing the geodesic dithers, we guarantee that during the optimization phase optimizing trajectories always lie on state manifolds. To analyze the behavior of the closed loop system, we employ averaging techniques developed for dynamical systems on Riemannian manifolds and apply results of closeness of solutions to obtain closeness of optimizing trajectories to local optimizers.

A recent version of extremum seeking algorithms for optimization of cost functions on submanifolds of Euclidean spaces appeared in [9]. In [9], the authors analyzed an extremum seeking algorithm based on the Lie bracket approach. The method presented in [9] is based upon embeddings of manifolds in Euclidean spaces and the techniques are inherited from extremum seeking algorithms in Euclidean spaces. However, in general, such embeddings may not be always available and implementation of extremum seeking algorithms on general Riemannian manifolds requires geometric extensions of methods developed in Euclidean spaces.

In terms of exposition, Section 2 presents some mathematical preliminaries needed for the analysis of the paper. Section 3 presents the extremum seeking problems on Riemannian manifolds and in Section 4 we extend the extremum seeking algorithm for optimization on Lie groups. In Sections 5 and 6 we present simple optimization examples on Lie groups  $SO(3)$  and  $SE(3)$  by applying the extremum seeking methods developed in Section 3.

**2. Preliminaries.** In this section we provide the differential geometric material which is necessary for the analysis presented in the rest of the paper. We define some of the frequently used symbols of this paper in Table 2.

**2.1. Riemannian manifolds.** DEFINITION 2.1 (see [24], Chapter 3). *A Riemannian manifold  $(M, g^M)$  is a differentiable manifold  $M$  together with a Riemannian metric  $g^M$  on  $M$ .*

TABLE 2.1  
Symbols and Their Descriptions

Symbol	Description
$M$	Riemannian manifold
$G$	Lie Group
$\star$	Lie group operation
$\mathfrak{X}(M)$	space of smooth time invariant vector fields on $M$
$\mathfrak{X}(M \times \mathbb{R})$	space of smooth time varying vector fields on $M$
$\nabla$	Levi-Civita connection
$T_x M$	tangent space at $x \in M$
$TM$	tangent bundle of $M$
$T_x^* M$	cotangent space at $x \in M$
$T^* M$	cotangent bundle of $M$
$\frac{\partial}{\partial x_i}$	basis tangent vectors at $x \in M$
$dx_i$	basis cotangent vectors at $x \in M$
$f(x, t)$	time varying vector fields on $M$
$\ \cdot\ _{g^M}$	Riemannian norm
$g^M(\cdot, \cdot)$	Riemannian metric on $M$
$d(\cdot, \cdot)$	Riemannian distance on $M$
$C^\infty(M)$	Space of smooth functions on $M$
$\Phi_f$	flow associated with $f$
$TF$	pushforward of $F$
$T_x F$	pushforward of $F$ at $x$
$\mathbb{R}_{>0}$	$(0, \infty)$
$\mathbb{R}_{\geq 0}$	$[0, \infty)$

nian metric  $g^M$ , where  $g^M$  is defined for each  $x \in M$  via an inner product  $g_x^M : T_x M \times T_x M \rightarrow \mathbb{R}$  on the tangent space  $T_x M$  (to  $M$  at  $x$ ) such that the function defined by  $x \mapsto g_x^M(X(x), Y(x))$  is smooth for any vector fields  $X, Y \in \mathfrak{X}(M)$ . In addition,

- (i)  $(M, g^M)$  is  $n$  dimensional if  $M$  is  $n$  dimensional;
- (ii)  $(M, g^M)$  is connected if for any  $x, y \in M$ , there exists a piecewise smooth curve that connects  $x$  to  $y$ .

Note that in the special case where  $M \doteq \mathbb{R}^n$ , the Riemannian metric  $g^M$  is defined everywhere by  $g_x^M \doteq \sum_{i=1}^n g_{ij}(x) dx_i \otimes dx_j$ ,  $x \in M$ , where  $\otimes$  is the tensor product on  $T_x^* M \times T_x^* M$  and  $g_{ij}$  is the  $(i, j)$  entity of  $g_x^M$ , see [24].

As formalized in Definition 2.1, connected Riemannian manifolds possess the property that any pair of points  $x, y \in M$  can be connected via a path  $\gamma \in \mathcal{P}(x, y)$ , where

$$(2.1) \quad \mathcal{P}(x, y) \doteq \left\{ \gamma : [a, b] \rightarrow M \left| \begin{array}{l} \gamma \text{ piecewise smooth,} \\ \gamma(a) = x, \gamma(b) = y \end{array} \right. \right\}.$$

**THEOREM 2.2** ([22], P. 94). *Suppose  $(M, g^M)$  is an  $n$  dimensional connected Riemannian manifold. Then, for any  $x, y \in M$ , there exists a piecewise smooth path  $\gamma \in \mathcal{P}(x, y)$  that connects  $x$  to  $y$ .*

The existence of connecting paths (via Theorem 2.2) between pairs of elements of an  $n$  dimensional connected Riemannian manifold  $(M, g^M)$  facilitates the definition

of a corresponding Riemannian distance. In particular, the Riemannian distance  $d : M \times M \rightarrow \mathbb{R}$  is defined by the infimal path length between any two elements of  $M$ , with

$$(2.2) \quad d(x, y) \doteq \inf_{\gamma \in \mathcal{P}(x, y)} \int_a^b \sqrt{g_{\gamma(t)}^M(\dot{\gamma}(t), \dot{\gamma}(t))} dt.$$

Note that since  $\mathcal{P}(x, y)$  contains piecewise smooth paths connecting  $x$  and  $y$  then  $\dot{\gamma}$  corresponds to left and right derivatives at non-smooth points of  $\gamma$ .

Using the definition of Riemannian distance  $d$  of (2.2), it may be shown that  $(M, d)$  defines a metric space, see [22]. Next, the crucial concept of *pushforward* operators is introduced.

**DEFINITION 2.3.** *For a given smooth mapping  $F : M \rightarrow N$  from manifold  $M$  to manifold  $N$  the pushforward  $TF$  is defined as a generalization of the Jacobian of smooth maps between Euclidean spaces, with*

$$TF : TM \rightarrow TN,$$

where

$$T_x F : T_x M \rightarrow T_{F(x)} N,$$

and

$$T_x F(X_x) \circ h = X_x(h \circ F), \quad x \in M, X_x \in T_x M, h \in C^\infty(N).$$

**2.2. Geodesic Curves.** Geodesics are defined [16] as length minimizing curves on Riemannian manifolds which satisfy

$$\nabla_{\dot{\gamma}(t)} \dot{\gamma}(t) = 0,$$

where  $\gamma(\cdot)$  is a geodesic curve on  $(M, g^M)$  and  $\nabla$  is the *Levi-Civita* connection on  $M$ , see [22]. The solution of the Euler-Lagrange variational problem associated with the length minimizing problem shows that all the geodesics on an  $n$  dimensional Riemannian manifold  $(M, g^M)$  must satisfy the following system of ordinary differential equations:

$$(2.3) \quad \ddot{\gamma}_i(s) + \sum_{j,k=1}^n \Gamma_{j,k}^i \dot{\gamma}_j(s) \dot{\gamma}_k(s) = 0, \quad i = 1, \dots, n,$$

where

$$(2.4) \quad \Gamma_{j,k}^i = \frac{1}{2} \sum_{l=1}^n g^{il} (g_{jl,k} + g_{kl,j} - g_{jk,l}), \quad g_{jl,k} = \frac{\partial g_{jl}}{\partial x_k},$$

where all the indices  $i, j, k, l$  run from 1 up to  $n = \dim(M)$  and  $[g^{ij}] \doteq [g_{ij}]^{-1}$ . Recall that  $g_{ij}$  is the  $(i, j)$  entity of the matrix  $g_x^M$ .

**DEFINITION 2.4** ([22], P. 72). *The restricted exponential map is defined by*

$$\exp_x : T_x M \rightarrow M, \quad \exp_x(v) = \gamma_v(1), v \in T_x M,$$

where  $\gamma_v(1)$  is the unique maximal geodesic [22, P. 59] initiating from  $x$  with the velocity  $v$  up to one as a solution of (2.3).

Throughout, *restricted exponential maps* are referred to as *exponential maps*. An open ball of radius  $\delta > 0$  and centered at  $0 \in T_x M$  in the tangent space at  $x$  is denoted by  $B_\delta(0) \doteq \{v \in T_x M \mid \|v\|_{g^M} < \delta\}$ . Similarly, the corresponding closed ball is denoted by  $\overline{B}_\delta(0)$ . Using the local diffeomorphic property of exponential maps, the corresponding geodesic ball centered at  $x$  is defined as follows.

DEFINITION 2.5. *For a vector space  $V$ , a star-shaped neighborhood of  $0 \in V$  is any open set  $U$  such that if  $u \in U$  then  $\alpha u \in U, \alpha \in [0, 1]$ .*

DEFINITION 2.6 ([22], p. 76). *A normal neighborhood around  $x \in M$  is any open neighborhood of  $x$  which is a diffeomorphic image of a star shaped neighborhood of  $0 \in T_x M$  under the exponential map  $\exp_x$ .*

LEMMA 2.7 ([22], Lemma 5.10). *For any  $x \in M$ , there exists a neighborhood  $B_\delta(0)$  in  $T_x M$  on which  $\exp_x$  is a diffeomorphism onto  $\exp_x(B_\delta(0)) \subset M$ .*

DEFINITION 2.8 ([22], p. 76). *In a neighborhood of  $x \in M$ , where  $\exp_x$  is a local diffeomorphism (this neighborhood always exists by Lemma 2.7), a geodesic ball of radius  $\delta > 0$  is denoted by  $\exp_x(B_\delta(0)) \subset M$ . The corresponding closed geodesic ball is denoted by  $\exp_x(\overline{B}_\delta(0))$ .*

DEFINITION 2.9. *The injectivity radius of  $M$  is*

$$\iota(M) \doteq \inf_{x \in M} \iota(x),$$

where

$$\iota(x) \doteq \sup\{r \in \mathbb{R}_{\geq 0} \mid \exp_x \text{ is diffeomorphic onto } \exp_x(B_r(0))\}.$$

DEFINITION 2.10. *The metric ball with respect to  $d$  on  $(M, g^M)$  is defined by*

$$B(x, r) \doteq \{y \in M \mid d(x, y) < r\}.$$

The following lemma reveals a relationship between normal neighborhoods and metric balls on  $(M, g^M)$ .

LEMMA 2.11 ([31], p. 122). *Given any  $\epsilon \in \mathbb{R}_{>0}$  and  $x \in M$ , suppose that  $\exp_x$  is a diffeomorphism on  $B_\epsilon(0) \subset T_x M$ . If  $B(x, r) \subset \exp_x B_\epsilon(0)$  for some  $r \in \mathbb{R}_{>0}$ , then*

$$\exp_x B_r(0) = B(x, r).$$

We note that  $B_\epsilon(0)$  is the metric ball of radius  $\epsilon$  with respect to the Riemannian metric  $g^M$  in  $T_x M$ . This paper focuses on dynamical systems governed by differential equations on a connected  $n$  dimensional Riemannian manifold  $(M, g^M)$ . Locally these differential equations are defined by (see [24])

$$\begin{aligned} \dot{x}(t) &= f(x(t), t), \quad f \in \mathfrak{X}(M \times \mathbb{R}), \\ x(0) &= x_0 \in M, t \in [t_0, t_f] \subset \mathbb{R}. \end{aligned}$$

The time dependent flow associated with a differentiable time dependent vector field  $f$  is a map  $\Phi_f$  satisfying

$$(2.5) \quad \begin{aligned} \Phi_f &: [t_0, t_f] \times [t_0, t_f] \times M \rightarrow M, \\ (s_f, s_0, x) &\mapsto \Phi_f(s_f, s_0, x) \in M, \end{aligned}$$

and

$$\frac{d\Phi_f(s, s_0, x)}{ds} \Big|_{s=t} = f(\Phi_f(t, s_0, x), t).$$

One may show, for a smooth vector field  $f$ , that the integral flow  $\Phi_f(s, t_0, \cdot) : M \rightarrow M$  is a local diffeomorphism, see [24].

**2.3. Lie groups.** As is well-known a Lie group  $(G, \star)$  is a Riemannian manifold equipped with operations  $g_1 \star g_2$  and  $g^{-1}$  which are smooth in their topologies ( $\star$  is the group operation of  $G$ ), see [20, 42]. We recall that the Lie algebra  $\mathcal{L}$  of a Lie group  $G$  (see [7], [42]) is the tangent space at the identity element  $e$  with the associated Lie bracket defined on the tangent space  $\mathcal{L} \doteq T_e G$  of  $G$ . A vector field  $X$  on  $G$  is called *left invariant* if

$$\forall g_1, g_2 \in G, \quad X(g_1 \star g_2) = T_{g_2} g_1 X(g_2),$$

where  $g \star : G \rightarrow G$ ,  $g \star (h) = g \star h$ ,  $T_{g_2} g : T_{g_2} G \rightarrow T_{g \star g_2} G$  which immediately implies  $X(g \star e) = X(g) = T_e L_g X(e)$ . For a left invariant vector field  $X$ , we define the *exponential map* on Lie groups as follows:

$$(2.6) \quad \exp : \mathcal{L} \rightarrow G, \quad \exp(tX(e)) \doteq \Phi(t, X), t \in \mathbb{R},$$

where  $\Phi(t, X)$  is the solution of  $\dot{g}(t) = X(g(t))$  with the boundary condition  $g(0) = e$ . It may be shown that the solution of  $\dot{g}(t) = X(g(t))$  with initial condition  $g_0 \in G$  is given by  $g_0 \star \exp(tX(e))$  where  $\star$  is the group operation of  $G$ , see [7]. A Riemannian metric  $g^G$  on a Lie group  $G$  is left invariant if

$$g_{g_2}^G(X(g_2), Y(g_2)) = g_{g_1 \star g_2}^G(T_{g_2} g_1 X(g_2), T_{g_2} g_1 Y(g_2)), \quad X, Y \in \mathfrak{X}(G).$$

Analogous to left invariant metrics, right invariant Riemannian metrics on  $G$  are defined. A Riemannian metric which is both left and right invariant is called *bi-invariant*. The Levi-Civita connection corresponding to a left invariant Riemannian metric  $g^G$  is denoted by  $\nabla^G$ . It may be shown that the Levi-Civita connection of a left invariant metric is left invariant i.e. (see [7])

$$Tg(\nabla_X^G Y) = \nabla_{TgX}^G TgY.$$

Note that the pushforward  $Tg$  is not evaluated at any point, hence,  $TgX$  is a well defined vector field on  $G$ .

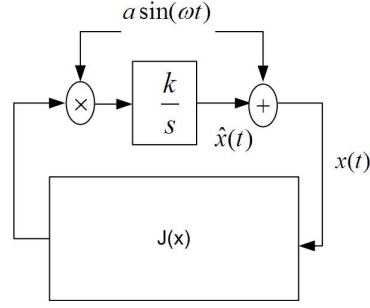
The following lemma gives a relationship between the exponential maps (2.6) and geodesics on Lie groups.

**LEMMA 2.12 ([30]).** *Assume  $G$  is a Lie group which admits a bi-invariant Riemannian metric. Then, for a left invariant vector field  $X$  on  $G$  we have*

$$\exp(tX(e)) = \exp_e(tX),$$

where  $\exp(tX(e))$  is the exponential map (2.6) and  $\exp_e(tX)$  is the geodesic emanating from  $e$  by velocity  $X(e)$ .

Note that  $\exp(tX(e))$  is the solution of the left invariant vector field  $X$  on  $G$ , whereas  $\exp_e(tX)$  is the solution of (2.3) with the initial conditions  $\exp_e(0X) = e$ ,  $\frac{d}{dt} \exp_e(tX)|_{t=0} = X(e)$ .

FIG. 3.1. Extremum seeking closed loop system for extremizing  $J : M \rightarrow \mathbb{R}_{\geq 0}$ ,  $M = \mathbb{R}$ .

**3. Optimization on Manifolds and Extremum Seeking.** Let us consider the optimization of a smooth function  $J : M \rightarrow \mathbb{R}_{\geq 0}$ , where  $(M, g^M)$  is an  $n$  dimensional smooth Riemannian manifold. Extremum seeking algorithms are a class of online optimization methods developed for minimizing/maximizing smooth functions defined on Euclidean spaces. These methods can be applied to both static and dynamic functions. In this paper we restrict our analysis to static functions defined on Riemannian manifolds. An extremum seeking closed loop for  $M = \mathbb{R}$  is shown in Figure 3.1. This is the simplest form of the extremum seeking algorithm to minimize/maximize a scalar function  $J : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ , see [38]. The dither signal  $a \sin(\omega t)$  provides a variation of the searching signal  $\hat{x}(t)$  in the one dimensional space  $\mathbb{R}$ . The output  $x(t) \in \mathbb{R}$  of the extremum seeking controller at time  $t$  is  $x(t) = \hat{x}(t) + a \sin(\omega t)$ , where  $\hat{x}(t) \in \mathbb{R}$  is the corresponding output of the integrator shown. The closed loop dynamics in  $\hat{x}$  coordinates are described by

$$(3.1) \quad \dot{\hat{x}}(t) = ka \sin(\omega t) J(\hat{x}(t) + a \sin(\omega t)).$$

The next lemma shows that, on average, the extremum seeking scheme of Figure 3.1 is a perturbation of the gradient algorithm.

**LEMMA 3.1** ([38]). *Consider the extremum seeking scheme in Figure 3.1. Then, on average, the closed loop of the extremum seeking algorithm in Figure 3.1 is a perturbation of the gradient algorithm in  $\hat{x}$  coordinates.*

The proof is based on the Taylor expansion of the cost function  $J$  in  $\hat{x}$  coordinates. By fixing  $\hat{x}(t)$  to a dummy variable  $z$ , we have

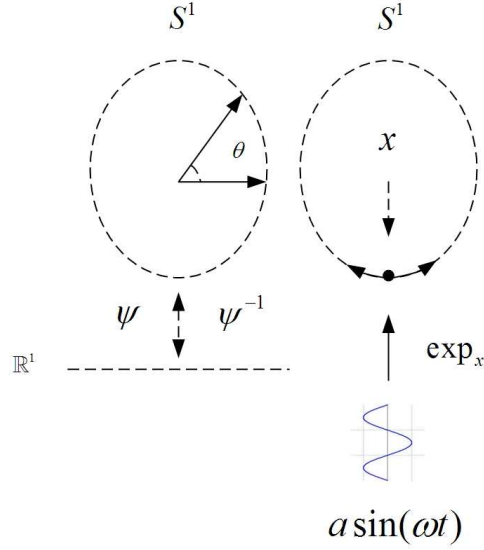
$$J(z + a \sin(\omega t)) = J(z) + \frac{\partial J}{\partial \hat{x}}|_{\hat{x}=z} a \sin(\omega t) + O(a^2).$$

Hence, the dynamical equations in  $\hat{x}$  coordinates are given by

$$(3.2) \quad \dot{\hat{x}}(t) = ka \sin(\omega t) \left( J(\hat{x}) + \frac{\partial J}{\partial \hat{x}}|_{\hat{x}=\hat{x}(t)} a \sin(\omega t) + O(a^2) \right).$$

The dynamical equation (3.2) is a periodic time-varying system where one may apply the averaging techniques, see [17]. In particular, the averaged system is given by

$$(3.3) \quad \begin{aligned} \dot{\hat{x}}(t) &= \frac{1}{T} \int_0^T ka \sin(\omega t) \left( J(\hat{x}) + \frac{\partial J}{\partial \hat{x}}|_{\hat{x}=\hat{x}(t)} a \sin(\omega t) + O(a^2) \right) dt \\ &= \frac{ka^2}{2} \frac{\partial J}{\partial \hat{x}}|_{\hat{x}=\hat{x}(t)} + O(a^4), \end{aligned}$$

FIG. 3.2. Geodesic dither at  $x$  on  $S^1$ .

where  $T$  is the period of  $\sin(\omega t)$ . Obviously (3.3) is in a perturbation form of the gradient algorithm in  $\mathbb{R}$ . The results of Lemma 3.1 are of great importance since convergence of the averaged dynamical system in (3.3) to a neighborhood of an equilibrium (local minimum or maximum) is required in order to guarantee the closeness of solutions of the time varying system (3.2) to a local optimizer, see [21]. Here, we extend the framework above for optimization of cost functions defined on finite dimensional Riemannian manifolds. The main challenge is to introduce a class of dither signals which perturb the optimizer  $\hat{x}$  without violating the restrictions imposed by the ambient manifolds. This is done by employing the so-called *geodesic dithers* as follows.

Consider an  $n$  dimensional Riemannian manifold  $(M, g^M)$ . For any  $x \in M$ , we consider the following local time-varying perturbation

$$(3.4) \quad x_p(t) = \exp_x \left( \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i} \right), \quad 0 < a_i,$$

where  $\{\frac{\partial}{\partial x_i}\}$ ,  $i = 1, \dots, n$ , is the basis for the tangent space at  $x$ . As formalized in Definition 2.4,  $\exp_x v$  is a geodesic emanating from  $x \in M$  with velocity  $v \in T_x M$ . In this case we perturb the  $n$  different coordinates on  $M$  with different frequencies  $\omega_i$ ,  $i = 1, \dots, n$ . As an example, for a one dimensional Riemannian manifold  $S^1$ , Figure 3.2 shows the example of a geodesic dither  $\exp_x a \sin(\omega t) \frac{\partial}{\partial \theta}$  at  $x \in S^1$  where a local coordinate system corresponding to  $S^1$  is given by (for the definition of coordinate systems see [23])

$$\psi : \theta \rightarrow (\sin(\theta), \cos(\theta)) \in \mathbb{R}^2, \quad \theta \in (0, 2\pi) \subset \mathbb{R}.$$

Motivated by the classical extremum seeking closed loop of Figure 3.1, we present a



time-varying extremum seeking vector field  $f(\hat{x}, t) \in T_{\hat{x}}M$  for optimization on  $(M, g^M)$  which is locally given by

$$(3.5) \quad f(\hat{x}, t) \doteq k \sum_{i=1}^n a_i \sin(\omega_i t) J \left( \exp_{\hat{x}} \sum_{j=1}^n a_j \sin(\omega_j t) \frac{\partial}{\partial x_j} \right) \frac{\partial}{\partial x_i}.$$

In this paper we assume that the optimization problem is to minimize a cost function, hence, following (3.3), without loss of generality assume  $k = -1$ . Finally, the optimizing trajectory  $\hat{x}(\cdot)$  is a solution of the time dependent differential equation

$$(3.6) \quad \dot{\hat{x}}(t) = f(\hat{x}(t), t) \in T_{\hat{x}(t)}M.$$

The closed loop system (3.6) is called the *extremum seeking system* on  $(M, g^M)$ . Note that  $t$  appears as a parameter in  $\exp_{\hat{x}} \left( \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i} \right)$ . That is to say

$$\exp_{\hat{x}} \left( \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i} \right) = \exp_{\hat{x}} \left( \eta \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i} \right) \Big|_{\eta=1},$$

where  $\eta$  and  $t$  are independent. Also note that the optimization algorithm (3.5) does not require any information about the gradient of  $J$ . However, cost function  $J$  should be measurable for the optimizing algorithm.

The next lemma proves that on compact Riemannian manifolds one may choose parameters  $a_i > 0$  sufficiently small such that for all  $x \in M$  we have

$\exp_x \left( \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i} \right) \in \iota(x)$ . These results will be employed to obtain the Taylor expansion of cost functions on Riemannian manifolds along geodesics.

**LEMMA 3.2.** *Consider the geodesic dither introduced by (3.4) on a smooth  $n$  dimensional compact Riemannian manifold  $(M, g^M)$ . Then for all  $x \in M$ , we may select  $a_i > 0$ ,  $i = 1, \dots, n$ , such that for all  $t \in \mathbb{R}$ ,  $\exp_x \left( \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i} \right) \in \iota(x)$ .*

*Proof.* Since  $M$  is compact and smooth then  $\iota(M)$  is bounded from below, see [19]. Hence, for all  $x \in M$  there exists  $\kappa \in \mathbb{R}_{>0}$  such that  $\kappa < \iota(x)$ . The Riemannian norm of the dither signal is given by

$$\left\| \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i} \right\|_{g^M}^2 = \sum_{i,j=1}^n a_i a_j \sin(\omega_i t) \sin(\omega_j t) g^M \left( \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right).$$

Since  $M$  is compact,  $\sum_{i,j=1}^n g^M \left( \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right)$  attains its maximum on  $M$ . Hence,  $a_i > 0$ ,  $i = 1, \dots, n$ , may be selected sufficiently small such that  $\| \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i} \|_{g^M}^2 \leq \kappa^2$ ,  $\forall t \in \mathbb{R}$ , which completes the proof.  $\square$

We adopt the following assumption for the cost function  $J$  on  $(M, g^M)$  and the dither frequencies  $\omega_i$ ,  $i = 1, \dots, n$ . This assumption is compatible with the main assumption on dither frequencies in [11] for multi-agent extremum seeking algorithms.

**ASSUMPTION 1.** *Cost function  $J : M \rightarrow \mathbb{R}_{\geq 0}$  is smooth and locally positive definite in a neighborhood of a unique local minimum  $x^* \in M$ , where  $J(x^*) = 0$ . The dither frequencies are  $\omega_i = m\bar{\omega}_i$ , where  $\bar{\omega}_i$  is rational,  $\bar{\omega}_i \neq \bar{\omega}_j$ ,  $2\bar{\omega}_i \neq \bar{\omega}_j$ ,  $j \neq i$  and  $\bar{\omega}_i \neq \bar{\omega}_j + \bar{\omega}_k$  for distinct  $i, j, k$ , where  $\omega_i, m \in \mathbb{R}_{>0}$ ,  $i, j \in 1, \dots, n$ .*

Here we introduce the *gradient and average systems* which correspond to the extremum seeking vector field (3.5) on  $(M, g^M)$ . For the smooth function  $J : M \rightarrow$

$\mathbb{R}_{\geq 0}$  the gradient system is defined by

$$(3.7) \quad \dot{x}(t) = - \sum_{i=1}^n \nabla_{\frac{\partial}{\partial x_i}} J(x(t)) \frac{\partial}{\partial x_i},$$

where the set  $\{\frac{\partial}{\partial x_i}, i = 1, \dots, n\}$  is a basis of  $T_x M$ . We note that the formal definition of the gradient  $\text{grad}J$  of  $J$  is given by [16] as

$$(3.8) \quad dJ(X) = g^M(\text{grad}J, X), \quad X \in \mathfrak{X}(M),$$

where  $dJ$  is the one form differential of  $J$  locally given by  $dJ = \sum_{i=1}^n \frac{\partial J}{\partial x_i} dx_i \in T_x^* M$ . Note that the existence of  $\text{grad}J$  in (3.8) is implied by an application of Riesz representation theorem since  $dJ$  belongs to the dual space  $T_x^* M$  and  $g^M(\cdot, \cdot)$  defines an inner product on  $T_x M$ , see [33]. In this case

$$\text{grad}J(x) = \sum_{i,j=1}^n g^{ij}(x) \nabla_{\frac{\partial}{\partial x_i}} J(x) \frac{\partial}{\partial x_j},$$

where  $[g^{ij}] = [g_{ij}]^{-1}$ . Hence, the formal gradient system corresponding to  $J$  is  $\dot{x}(t) = - \sum_{i=1}^n g^{ij}(x(t)) \nabla_{\frac{\partial}{\partial x_i}} J(x(t)) \frac{\partial}{\partial x_j}$ . However, in this paper, we adopt the terminology that the gradient system of  $J$  refers to (3.7). The *scaled* version of the gradient system (3.7) is given as

$$(3.9) \quad \dot{x}(t) = \hat{f}(x(t)) \doteq - \sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x(t)) \frac{\partial}{\partial x_i}, a_i > 0.$$

With no further confusion we refer the scaled gradient system as gradient system. For the periodic time varying vector field  $f(x, t)$  in (3.5), the averaged dynamical system is defined as follows

$$(3.10) \quad \dot{x}(t) = \hat{f}(x(t)) \doteq \frac{1}{T} \int_0^T f(x(t), \tau) d\tau,$$

where  $T$  is the period of  $f$ , i.e.  $f(x, t) = f(x, t + T)$ .

The following lemma proves that, on average, the closed loop of the extremum seeking system (3.6) is a perturbation of the gradient system of the cost function  $J$ .

**LEMMA 3.3.** *Consider the extremum seeking system in (3.6) on a compact Riemannian manifold  $(M, g^M)$  where the optimizing trajectory is perturbed by the geodesic dither presented in (3.4). Then, subject to Assumption 1, the averaged dynamical system of (3.5) is a perturbation of the gradient system (3.9) of the cost function  $J$ .*

*Proof.* See Appendix A.  $\square$

The results of Lemma 3.3 imply that the state trajectories of the averaged dynamical system (3.10) can be estimated by the state trajectories of the scaled gradient system (3.9). In the case  $a_i = a_j$ ,  $i, j = 1, \dots, n$  then  $\dot{x}(t) = - \sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x(t)) \frac{\partial}{\partial x_i}$  is identical to (3.7).

**REMARK 1.** *Note that the compactness of  $M$  can be relaxed when the analysis is carried out in a local neighborhood of  $x^*$  which is contained in a compact set. The existence of this compact set is guaranteed since manifolds are Hausdorff spaces and Hausdorff spaces are locally compact, see [23], Proposition 4.27.*

We analyze the properties of the state trajectory of (3.5) based on the state trajectory of the average system (3.10). Also stability properties of the gradient system  $\dot{x}(t) = -\sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x(t)) \frac{\partial}{\partial x_i}$  facilitate the closeness of solutions between the time varying dynamical systems (3.5) and the gradient system (3.9). The same results on Euclidean spaces are presented in [34]. The following lemma gives the uniform local asymptotic stability of the gradient system obtained in the proof of Lemma 3.3.

**LEMMA 3.4.** *Consider the gradient dynamical system (3.9) on a compact connected  $n$  dimensional Riemannian manifold  $(M, g^M)$ . Then, subject to Assumption 1, the cost function  $J$  is a Lyapunov function and  $x^*$  is locally asymptotically stable (see [17]) for the gradient system (3.9) on  $(M, g^M)$*

*Proof.* See Appendix B.  $\square$

The following theorem is the main result of this paper which gives a local convergence of the geodesic extremum seeking system to a unique local minimum/maximum of the function  $J$  on an  $n$  dimensional compact Riemannian manifold  $(M, g^M)$ . The compactness assumption can be relaxed if the analysis is restricted to a local neighborhood of the optimizer  $x^*$  which is contained in a compact set as per Remark 1.

**THEOREM 3.5.** *Consider the geodesic extremum seeking system (3.6) on a compact connected  $n$  dimensional Riemannian manifold  $(M, g^M)$ , where  $\omega_i = \omega \bar{\omega}_i, i = 1, \dots, n$  satisfy Assumption 1. Assume  $x^* \in M$  is a unique local optimizer of  $J : M \rightarrow \mathbb{R}_{\geq 0}$ , where  $J$  satisfies Assumption 1. Then for any neighborhood  $U_{x^*} \subset M$  of  $x^*$  on  $M$ , there exist sufficiently small parameters  $a_i > 0, i = 1, \dots, n$ , sufficiently large frequency  $\omega$  and a neighborhood of  $x^*$  denoted by  $\tilde{U}_{x^*} \subset M$ , such that for any  $x_0 \in \tilde{U}_{x^*}$ , the state trajectory of the closed loop system (3.5) initiating from  $x_0$  ultimately enters and remains in  $U_{x^*}$ .*

*Proof.* See Appendix C.  $\square$

**REMARK 2.** *The results of Theorem 3.5 give closeness of the state trajectory of the extremum seeking system (3.5) to the local optimizer  $x^*$  since the state trajectory of (3.5) ultimately enters and remains in  $U_{x^*}$ . This is guaranteed by tuning the frequency  $\omega$  and dither amplitudes  $a_i > 0$ .*

**REMARK 3.** *The proof of Theorem 3.5 is based on closeness of solutions and averaging analysis for dynamical systems evolving on Riemannian manifolds and stability of perturbed systems. The results related to averaging for dynamical systems on Riemannian manifolds are presented in Appendix D. The stability results for perturbed systems on Riemannian manifolds are presented in appendix F. We employ the averaging techniques presented in Appendix D to analyze the closeness of solutions on Riemannian manifolds where the averaged dynamical system is not necessarily stable. However, one may show that the state trajectory of the averaged system remains bounded in a neighborhood of  $x^*$ .*

The following lemma presents the state trajectory of a special combination of flows on  $(M, g^M)$ . The proof of Theorem 3.5 is based on closeness of solutions of the state trajectory of extremum seeking system (3.5) and the trajectory  $z(\cdot)$  constructed in the lemma below.

**DEFINITION 3.6.** *Let  $X, Y \in \mathfrak{X}(M)$  be smooth vector fields on  $M$ , where it may be shown that  $\Phi_Y(t, t_0, \cdot) : M \rightarrow M$  is a local diffeomorphism (see [1]). Let us denote  $T_x \Phi_Y^{(t, t_0)^{-1}}$  as the pushforward of  $\Phi_Y^{-1}(t, t_0, \cdot) : M \rightarrow M$  at  $x \in M$ . Define the pull*

back of  $\Phi_Y^{(t,t_0)} \doteq \Phi_Y(t, t_0, \cdot)$  denoted by  $\Phi_Y^{(t,t_0)*}$  as follows.

$$(3.11) \quad \begin{aligned} \Phi_Y^{(t,t_0)*} : \mathfrak{X}(M) &\rightarrow \mathfrak{X}(M), \\ \left( \Phi_Y^{(t,t_0)*} X \right) (x) &\doteq T_{\Phi_Y(t,t_0,x)} \Phi_Y^{(t,t_0)^{-1}} X(\Phi_Y(t, t_0, x)), \\ X &\in \mathfrak{X}(M), x \in M, t \in \mathbb{R}. \end{aligned}$$

In  $\mathbb{R}^n$ , for a diffeomorphism  $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $X \in \mathfrak{X}(\mathbb{R}^n)$ , we have

$$(\phi^* X)(x) = \left( \frac{\partial \phi^{-1}}{\partial x} \circ X \circ \phi \right)(x) = \frac{\partial \phi^{-1}}{\partial x}(X(\phi(x))).$$

LEMMA 3.7. Consider a  $T$  periodic time varying dynamical system  $\dot{x} = \epsilon f(x, t)$ , where  $f(x, t + T) = f(x, t)$ , on an  $n$  dimensional Riemannian manifold  $(M, g^M)$ . The averaged dynamical system is given by  $\dot{x} = \epsilon \hat{f}(x)$ , where  $\hat{f}(x) = \frac{1}{T} \int_0^T f(x, s) ds$  and  $\epsilon \in \mathbb{R}_{\geq 0}$ . Consider the combination of state flows  $z(t) \doteq \Phi_{\epsilon Z}^{(1,0)} \circ \Phi_{\epsilon f}(t, t_0, x_0) \doteq \Phi_{\epsilon Z}(1, 0, \Phi_{\epsilon f}(t, t_0, x_0)) \in M$  where  $Z(t, x) \doteq \int_0^t (\hat{f}(x) - f(x, s)) ds$ . Then  $z(\cdot)$  satisfies

$$\dot{z}(t) = \epsilon \left[ (\Phi^{-1})_{\epsilon Z}^{(1,0)*} f + \int_0^1 (\Phi^{-1})_{\epsilon Z}^{(1,s)*} (\hat{f} - f) ds \right] \circ z(t),$$

where  $(\Phi^{-1})_{\epsilon Z}^{(1,s)*}$  is the pullback of the local diffeomorphism  $(\Phi^{-1})_{\epsilon Z}^{(1,s)} \doteq \Phi_{\epsilon Z}^{-1}(1, s, \cdot)$ .

*Proof.* See Appendix D.  $\square$

The results of Lemma 3.7 are employed to obtain the closeness of solutions for the dynamical systems (3.5) and its averaged system as per Lemma 3.2. The full proof of Theorem 3.5 is given in Appendix C.

**4. Extremum seeking on Lie groups.** The extremum seeking system (3.6) is modified for optimization on Lie groups. In this case we employ the group structure of the ambient manifold and define the extremum seeking vector field along the exponential maps on Lie groups. This makes the computation of the geodesic dithers defined before particularly simple for matrix Lie groups. The following lemma characterizes geodesics on Lie groups which admit bi-invariant Riemannian metrics.

LEMMA 4.1. Assume  $G$  is a Lie group which admits a bi-invariant Riemannian metric then for a left invariant vector field  $X$  on  $G$ ,  $g \star \exp(tX(e))$  is a geodesic emanating from  $g \in G$ .

*Proof.* The proof is a straightforward extension of Lemma 2.12. To show that  $\gamma(t) = g \star \exp(tX)$ ,  $X \in \mathcal{L}$  is a geodesic through  $g \in G$  we have

$$\begin{aligned} \nabla_{\dot{\gamma}(t)}^G \dot{\gamma}(t) &= \nabla_{T_{\exp(tX)g} \frac{d \exp(tX)}{dt}}^G T_{\exp(tX)g} \frac{d \exp(tX)}{dt} \\ &= T_{\exp(tX)g} \nabla_{\frac{d \exp(tX)}{dt}}^G \frac{d \exp(tX)}{dt} = 0, \end{aligned}$$

since  $\exp(tX)$  is a geodesic by Lemma 2.12 and  $\nabla_{\frac{d \exp(tX)}{dt}}^G \frac{d \exp(tX)}{dt} = 0$   $\square$

The geodesic dither (3.4) is given along  $\exp$  on Lie groups by

$$(4.1) \quad g_p(t) = g \star \exp \left( \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial g_i} \right),$$

where  $\frac{\partial}{\partial g_i}$  are the base elements of  $\mathcal{L}$ . The extremum seeking vector field on  $G$  is then defined by

$$(4.2) \quad f(g, t) \doteq - \sum_{i=1}^n a_i \sin(\omega_i t) J \left( g \star \exp \left( \sum_{j=1}^n a_j \sin(\omega_j t) \frac{\partial}{\partial g_j} \right) \right) T_e g \left( \frac{\partial}{\partial g_i} \right) \in T_g G,$$

where  $T_e g \left( \frac{\partial}{\partial g_i} \right)$ ,  $i = 1, \dots, n$  are tangent vectors at  $T_g G$ . One may easily verify that  $T_e(g_1 \star g_2) \left( \frac{\partial}{\partial g_i} \right) = T_{g_2} g_1 \left( T_{g_2} \left( \frac{\partial}{\partial g_i} \right) \right)$  which shows that  $T_e g \left( \frac{\partial}{\partial g_i} \right)$  is a left invariant vector field. Note that  $f(g, t)$  is not necessarily left invariant since  $J(g_1 \star g_2 \star \exp(\sum_{i=1}^n a_i \sin(\omega_i t))) = J(g_2 \star \exp(\sum_{i=1}^n a_i \sin(\omega_i t)))$  is not true in general.

The following theorem gives the stability of the geodesic extremum seeking algorithm (4.2) on compact Lie groups.

**THEOREM 4.2.** *Consider the extremum seeking system (4.2) on a compact connected  $n$  dimensional Lie group  $G$ , where  $\omega_i = \omega \bar{\omega}_i$ ,  $i = 1, \dots, n$  satisfy Assumption 1. Assume  $g^* \in G$  is a unique local optimizer of  $J : G \rightarrow \mathbb{R}_{\geq 0}$ , where  $J$  satisfies Assumption 1. Then for any neighborhood  $U_{g^*} \subset G$  of  $G^*$  on  $G$ , there exist sufficiently small parameters  $a_i > 0$ ,  $i = 1, \dots, n$ , sufficiently large  $\omega$  and a neighborhood  $\hat{U}_{g^*} \subset G$  of  $g^*$  such that for any  $g_0 \in \hat{U}_{g^*}$ , the state trajectory of (4.2) initiating from  $g_0$  ultimately enters and remains in  $U_{g^*}$ .*

*Proof.* Since  $G$  is compact then it possesses a bi-invariant Riemannian metric. Hence, Lemma 4.1 implies that  $g \star \exp(\sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial g_i})$  is a geodesic through  $g \in G$  and the proof is identical to the proof of Theorem 3.5.  $\square$

In the case that  $G$  is not compact we employ the Taylor expansion of smooth functions on  $G$ , given in [20].

**LEMMA 4.3** ([20]). *Consider a left invariant vector field  $X \in \mathfrak{X}(G)$  which is identified by  $X(e) \in \mathcal{L}$ . Then for a smooth function  $J : G \rightarrow \mathbb{R}$  we have*

$$J(g \star \exp(X)) = \sum_{j=1}^m \frac{1}{j!} (X^j J)(g) + \frac{1}{m!} \int_0^1 (1-s)^m (X^{m+1} J)(g \star \exp(sX)) ds,$$

where  $X^j J(g) = \frac{d^j}{dt^j} J(g \star \exp(tX))|_{t=0}$ .

**LEMMA 4.4.** *Consider the extremum seeking algorithm in (4.2) on a Lie group  $G$  where the optimizing trajectory is perturbed by the exponential dither presented in (4.1). Then, subject to Assumption 1, the averaged dynamical system of (4.2) is a perturbation of the first order variation  $X_a(g)J$ , where  $X_a(g)$  is the left invariant vector field identified by  $X_a(e) = \sum_{i=1}^n a_i^2 \frac{\partial}{\partial g_i}$ .*

*Proof.* Parallel to the proof of Lemma 3.3 we have

$$\begin{aligned} J(g \star \exp \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial g_i}) &= J(g) + (XJ)(g) + \dots + \\ &\frac{1}{(m-1)!} \times (X^{m-1} J)(g) + \frac{1}{(m-1)!} \times \int_0^1 (1-s)^{m-1} X^m \\ &J \left( g \star \exp s \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial g_i} \right) ds, \end{aligned}$$

where  $X(g, t) = \sum_{i=1}^n a_i \sin(\omega_i t) T_e g \frac{\partial}{\partial g_i}$ . Following the proof of Lemma 3.3, after averaging and changing the time scale to  $\tau = \omega t$  we have

$$(4.3) \quad \frac{dg}{d\tau} = -\frac{1}{\omega} \sum_{i=1}^n \frac{a_i^2}{2} \left( T_e g \frac{\partial}{\partial g_i} J \right) (g) T_e g \frac{\partial}{\partial g_i} + \frac{1}{\omega} \sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) T_e g \frac{\partial}{\partial g_i},$$

which completes the proof.  $\square$

Note that since  $G$  is not compact the perturbation vector field  $\frac{1}{\omega} \sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) T_e g \frac{\partial}{\partial g_i}$  is not uniformly bounded on  $G$ . However, at any point  $g \in G$  its magnitude is of order  $O((\max_{i \in 1, \dots, n} a_i)^4)$ . It is shown in the proof of Theorem F.2 in Appendix F that selecting the perturbation of an asymptotic stable system on a Riemannian manifold sufficiently small guarantees that the state trajectory of the perturbed system remains in a compact neighborhood of the equilibrium of the asymptotic stable system. In that case the magnitude of perturbation vector field above can be uniformly bounded on a compact set containing the equilibrium.

LEMMA 4.5. *Consider the gradient dynamical system*

$$\dot{g} = -\sum_{i=1}^n \frac{a_i^2}{2} \left( T_e g \frac{\partial}{\partial g_i} J \right) (g) T_e g \frac{\partial}{\partial g_i}, \quad g \in G \text{ on an } n \text{ dimensional Lie group } (G, \star).$$

*Then subject to Assumption 1,  $g^*$  is locally asymptotically stable on  $(G, \star)$  for the gradient dynamical system.*

*Proof.* The proof parallels the proof of Lemma 3.4 since

$$(4.4) \quad \mathcal{L}_{-\sum_{i=1}^n \frac{a_i^2}{2} \left( T_e g \frac{\partial}{\partial g_i} J \right) (g) T_e g \frac{\partial}{\partial g_i}} J(g) = -\sum_{i=1}^n \frac{a_i^2}{2} \left( T_e g \frac{\partial}{\partial g_i} J \right)^2 (g) \leq 0.$$

$\square$

Note that by Assumption 1,  $(XJ)(g) = \lim_{t \rightarrow 0} \frac{J(g \star \exp(tX(e))) - J(g)}{t} = 0$  only at the unique local optimal point  $g^*$ .

THEOREM 4.6. *Consider the extremum seeking system (4.2) on a connected  $n$  dimensional Lie group  $G$ , where  $\omega_i = \omega \bar{\omega}_i$ ,  $i = 1, \dots, n$  satisfy Assumption 1. Assume  $g^* \in G$  is a unique local optimizer of  $J : G \rightarrow \mathbb{R}_{\geq 0}$ , where  $J$  satisfies Assumption 1. Then for any neighborhood  $U_{g^*} \subset G$  of  $G^*$  on  $G$ , there exist sufficiently small parameters  $a_i$ ,  $i = 1, \dots, n$ , sufficiently large  $\omega$  and a neighborhood  $\hat{U}_{g^*} \subset G$  of  $g^*$  such that for any  $g_0 \in \hat{U}_{g^*}$ , the state trajectory of (4.2) initiating from  $g_0$  ultimately enters and remains in  $U_{g^*}$ .*

*Proof.* See Appendix E.  $\square$

**5. Example on  $SO(3)$ .** In this section we give a conceptual example for orientation control which is modeled by elements of  $SO(3)$  which is a compact Lie group.

We recall that  $SO(3)$  is the rotation group in  $\mathbb{R}^3$  given by

$$SO(3) = \{g \in GL(3) \mid g \cdot g^T = I, \det(g) = 1\},$$

where  $GL(n)$  is the set of nonsingular  $n \times n$  matrices. The Lie algebra of  $SO(3)$  which is denoted by  $so(3)$  is given by (see [42])  $so(3) = \{X \in M(3) \mid X + X^T = 0\}$ , where  $M(n)$  is the space of all  $n \times n$  matrices. The Lie group operation  $\star$  is given by the matrix multiplication and consequently  $T_{g_1} g_2$  is also given by the matrix multiplication  $g_2 X$ ,  $X \in T_{g_1} G$ .

A left invariant dynamical system on  $SO(3)$  is given by

$$(5.1) \quad \dot{g}(t) = gX, \quad g(0) = g_0, \quad X \in so(3).$$

The Lie algebra bilinear operator is defined as the commuter of matrices, i.e.

$$[X, Y] = XY - YX, \quad X, Y \in so(3).$$

Equation (5.1) above is written as

$$\begin{pmatrix} \dot{g}_{11} & \dot{g}_{12} & \dot{g}_{13} \\ \dot{g}_{21} & \dot{g}_{22} & \dot{g}_{23} \\ \dot{g}_{31} & \dot{g}_{32} & \dot{g}_{33} \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{pmatrix} \begin{pmatrix} 0 & X_1(t) & X_3(t) \\ -X_1(t) & 0 & X_2(t) \\ -X_3(t) & -X_2(t) & 0 \end{pmatrix}.$$

The optimization is performed for the cost function  $J : SO(3) \rightarrow \mathbb{R}$  defined by

$$J(g) = \frac{1}{2} \text{tr}((g - g^*)^T (g - g^*)),$$

where  $g^*$  is the optimal orientation matrix in  $SO(3)$ . We assume  $g^* = I_{3 \times 3}$ , hence,

$J = 3 - \text{tr}(g)$ . The Lie algebra  $so(3)$  is spanned by  $\frac{\partial}{\partial g_1} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ ,  $\frac{\partial}{\partial g_2} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$  and  $\frac{\partial}{\partial g_3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$ . For this example the dither vector  $X(e)$  at the Lie algebra  $so(3)$  is given by

$$X(e) = \sum_{i=1}^3 a_i \sin(\omega_i t) \frac{\partial}{\partial g_i} = \begin{pmatrix} 0 & a_1 \sin(\omega_1 t) & a_3 \sin(\omega_3 t) \\ -a_1 \sin(\omega_1 t) & 0 & a_2 \sin(\omega_2 t) \\ -a_3 \sin(\omega_3 t) & -a_2 \sin(\omega_2 t) & 0 \end{pmatrix},$$

hence, the dither vector field is given by

$$X(g) = g \cdot \begin{pmatrix} 0 & a_1 \sin(\omega_1 t) & a_3 \sin(\omega_3 t) \\ -a_1 \sin(\omega_1 t) & 0 & a_2 \sin(\omega_2 t) \\ -a_3 \sin(\omega_3 t) & -a_2 \sin(\omega_2 t) & 0 \end{pmatrix},$$

where  $g \in SO(3)$ .

The extremum seeking vector field given in (4.2) is presented on  $SO(3)$  by

$$(5.2) \quad f(g, t) \doteq - \sum_{i=1}^3 a_i \sin(\omega_i t) J(g \exp \sum_{j=1}^3 a_j \sin(\omega_j t) \frac{\partial}{\partial g_j}) g \frac{\partial}{\partial g_i},$$

where by Lemma 2.12 the exponential map on  $SO(3)$  is a geodesic ( $SO(3)$  is compact).

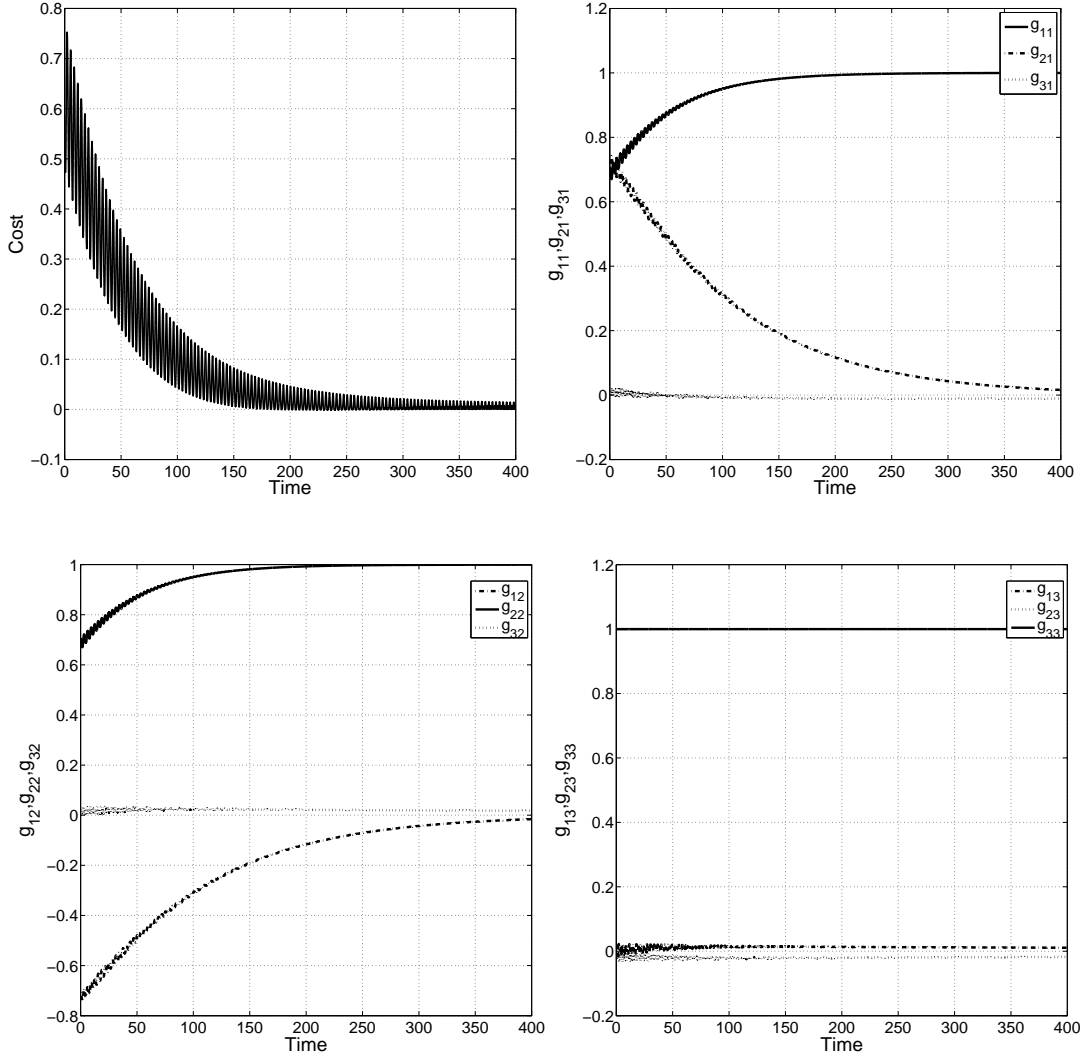
Note that in (5.2)  $g \frac{\partial}{\partial g_i} \in T_g SO(3)$ . The optimizing trajectory  $g(\cdot)$  is a solution of the time dependent differential equation

$$\dot{g}(t) = f(g, t) \in T_g SO(3).$$

The algorithms initiates from the initial orientation at  $g_0 = \begin{pmatrix} \cos(\frac{\pi}{4}) & -\sin(\frac{\pi}{4}) & 0 \\ \sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \in$

$SO(3)$ . The amplitudes and frequencies are set at  $a_1 = a_2 = a_3 = .1$  and  $\omega_1 = 2, \omega_2 = 4.1, \omega_3 = 6.2$ . Figure 5 shows the convergence of the cost function and the state tra-

jectory  $g(t) = \begin{pmatrix} g_{11}(t) & g_{12}(t) & g_{13}(t) \\ g_{21}(t) & g_{22}(t) & g_{23}(t) \\ g_{31}(t) & g_{32}(t) & g_{33}(t) \end{pmatrix} \in SO(3)$ .

FIG. 5.1. Cost and state convergence on  $SO(3)$ .

As is obvious, the optimal solution for the optimization problem is  $g^* = I_{3 \times 3}$ .

The algorithm converges to  $g = \begin{pmatrix} .999 & -0.0159 & 0.0110 \\ 0.0161 & 0.9998 & -0.0174 \\ -0.0108 & 0.0175 & 0.9998 \end{pmatrix} \in SO(3)$ .

**6. Example on  $SE(3)$ .** In this section we give another conceptual example for an orientation control on  $SE(3)$  which is not compact.

As is known,  $SE(3)$  is the space of rotation and translation which is used for robotic modeling. We have

$$SE(3) = \left\{ \begin{pmatrix} g_{SO(3)} & g_{\mathbb{R}} \\ 0_{1 \times 3} & 1 \end{pmatrix} \in \mathbb{R}^{4 \times 4} \mid g_{SO(3)} \in SO(3), g_{\mathbb{R}} \in \mathbb{R}^{3 \times 1} \right\},$$



where  $g_{SO(3)}$  models the rotation and  $g_{\mathbb{R}}$  models the translation in  $\mathbb{R}^3$ . The Lie algebra of  $SE(3)$  which is denoted by  $se(3)$  is given by (see [42])

$$se(3) = \left\{ \begin{pmatrix} S & v \\ 0_{1 \times 3} & 0 \end{pmatrix} \in \mathbb{R}^{4 \times 4} \mid S \in so(3), v \in \mathbb{R}^3 \right\},$$

Let us consider the cost function  $J : SE(3) \rightarrow \mathbb{R}$  as

$$J(g) = \frac{1}{2} \text{tr}((g_{SO(3)} - g_{SO(3)}^*)^T (g_{SO(3)} - g_{SO(3)}^*)) + \frac{1}{2} \|g_{\mathbb{R}} - r^*\|_{\mathbb{R}^3}^2,$$

where  $g_{SO(3)}^*$  is the optimal orientation matrix in  $SO(3)$  and  $r^*$  is the optimal distance from the origin in  $\mathbb{R}^3$ . As is obvious the optimal solution for the optimization problem above is  $\begin{pmatrix} g_{SO(3)}^* & r^* \\ 0_{1 \times 3} & 1 \end{pmatrix} \in SE(3)$ . The cost function above minimizes the distance from the orientation  $g_{SO(3)}^*$  and distance from  $r^*$ . Without loss of generality, we assume  $g_{SO(3)}^* = I_{3 \times 3} \in SO(3)$  and  $r^* = (0, 0, 0) \in \mathbb{R}^3$ .

The Lie algebra  $se(3)$  is spanned by

$$\begin{aligned} \frac{\partial}{\partial g_1} &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \frac{\partial}{\partial g_2} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \frac{\partial}{\partial g_3} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \frac{\partial}{\partial g_4} = \\ &\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \frac{\partial}{\partial g_5} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ and } \frac{\partial}{\partial g_6} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

For this example the dither vector  $X(e)$  at the Lie algebra  $se(3)$  is given by

$$X(e) = \sum_{i=1}^6 a_i \sin(\omega_i t) \frac{\partial}{\partial g_i} = \begin{pmatrix} 0 & a_1 \sin(\omega_1 t) & a_3 \sin(\omega_3 t) & a_4 \sin(\omega_4 t) \\ -a_1 \sin(\omega_1 t) & 0 & a_2 \sin(\omega_2 t) & a_5 \sin(\omega_5 t) \\ -a_3 \sin(\omega_3 t) & -a_2 \sin(\omega_2 t) & 0 & a_6 \sin(\omega_6 t) \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad (6.1)$$

hence, the dither vector field is given by  $X(g) = g \cdot X(e)$ , where  $g \in SE(3)$ .

Similar to the example on  $SO(3)$ , the extremum seeking vector field on  $SE(3)$  is given by the following vector field

$$f(g, t) = - \sum_{i=1}^6 a_i \sin(\omega_i t) J(g \exp \sum_{j=1}^6 a_j \sin(\omega_j t) \frac{\partial}{\partial g_j}) g \frac{\partial}{\partial g_i},$$

where  $\exp$  is the exponential operator defined on  $SE(3)$ . On  $SE(3)$  the exponential map does not coincide with geodesics since  $SE(3)$  does not admit a bi-invariant Riemannian metric, see [30]. Hence, the results of Theorem 4.6 grants the local convergence of the algorithm.

In this case, the  $\exp$  operator is not the same as the  $\exp$  operator on  $SO(3)$ .

For a tangent vector  $\begin{pmatrix} S & v \\ 0_{1 \times 3} & 0 \end{pmatrix} \in se(3)$ , where  $S = \begin{pmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{pmatrix}$ , we have

$$\exp\left(\begin{pmatrix} S & v \\ 0_{1 \times 3} & 0 \end{pmatrix}\right) = \begin{pmatrix} \exp(S) & Av \\ 0_{1 \times 3} & 1 \end{pmatrix}, \text{ where } A = I_{3 \times 3} + \frac{(1 - \cos(\theta))}{\theta^2} S + \frac{(\theta - \sin(\theta))}{\theta^3} S^2,$$

and  $\theta = \sqrt{a^2 + b^2 + c^2}$ . In the case that  $\theta = 0$ , we have  $\exp\left(\begin{pmatrix} S & v \\ 0_{1 \times 3} & 0 \end{pmatrix}\right) = \begin{pmatrix} \exp(S) & v \\ 0_{1 \times 3} & 1 \end{pmatrix}$ .

The optimizing trajectory  $g(\cdot)$  is a solution of the time dependent differential equation

$$\dot{g}(t) = f(g, t) \in T_g SE(3).$$

The algorithm initiates from the initial orientation at

$g_{SO(3)}(0) = \begin{pmatrix} \cos(\frac{\pi}{4}) - \sin(\frac{\pi}{4}) & 0 \\ \sin(\frac{\pi}{4}) & \cos(\frac{\pi}{4}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \in SO(3)$  and  $g_{\mathbb{R}} = (1, -1, 2) \in \mathbb{R}^3$ . The amplitudes and frequencies are set at  $a_1 = \dots = a_6 = .1$  and  $\omega_i = 2i + \epsilon_i$ ,  $i = 1, \dots, 6$ .

The results for the convergence of the cost function and the state trajectory  $g(t) = \begin{pmatrix} g_{11}(t) & g_{12}(t) & g_{13}(t) & g_{14}(t) \\ g_{21}(t) & g_{22}(t) & g_{23}(t) & g_{24}(t) \\ g_{31}(t) & g_{32}(t) & g_{33}(t) & g_{34}(t) \\ 0 & 0 & 0 & 1 \end{pmatrix} \in SE(3)$  are shown in Figures 6.1 and 6.2.

**7. Conclusion.** In this paper we extended the standard extremum seeking algorithms developed for online optimization to a class of online algorithms for optimization on Riemannian manifolds. We introduced the notion of geodesic dithers for extremum seeking algorithms on Riemannian manifolds and employed the results of averaging on manifolds to obtain a local convergence of the extremum seeking loop to a local optimizer on Riemannian manifolds. Two examples on Lie groups were presented to illustrate the efficacy of the proposed algorithm.

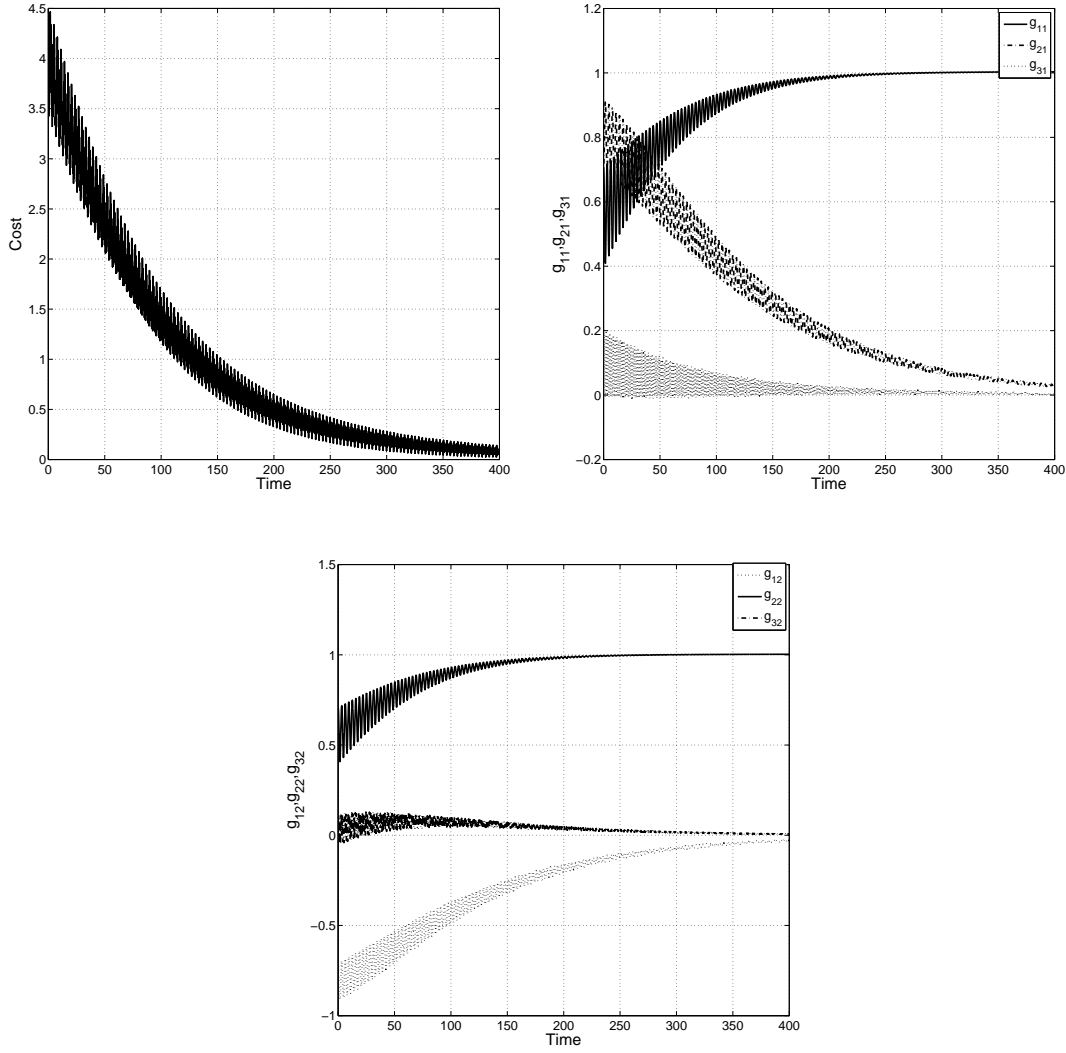
**Appendix A. Proof of Lemma 3.3 .** The cost function  $J$  may be expanded along geodesics by using the Taylor expansion on Riemannian manifolds, see [36]. We employ Lemma 3.1 to guarantee that there exist  $a_i > 0$ ,  $i = 1, \dots, n$ , such that  $\exp_x\left(\sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i}\right) \in \iota(x)$ . Then the Taylor expansion of  $J$  at  $x \in M$  along the geodesic  $\exp_x(\eta X)$ , where  $X \in T_x M$ , is given by (see [36])

$$\begin{aligned} J(\exp_x \eta X) &= J(x) + \eta(\nabla_X J)(x) + \dots + \frac{\eta^{m-1}}{(m-1)!} \times \\ (A.1) \quad &(\nabla_X^{m-1} J)(x) + \frac{\eta^m}{(m-1)!} \int_0^1 (1-s)^{m-1} \nabla_X^m J(\exp_x(s\eta X)) ds, \quad 0 < \eta < \eta^*, \end{aligned}$$

which is equivalent to

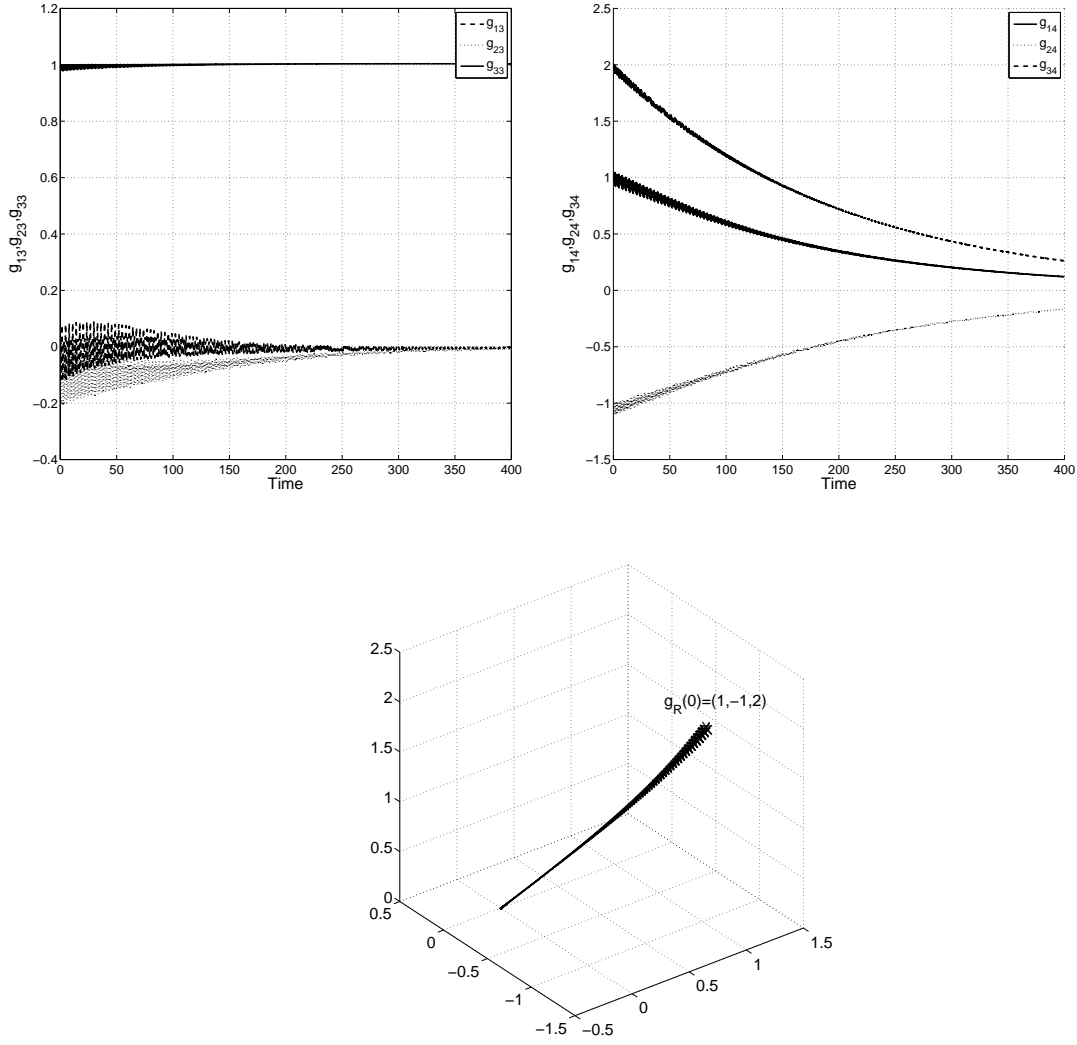
$$\begin{aligned} J(\exp_x \eta X) &= J(x) + \eta(dJ(X))|_x + \dots + \frac{\eta^{m-1}}{(m-1)!} \times (\nabla_X^{m-2} dJ)(X)|_x + \\ (A.2) \quad &\frac{\eta^m}{(m-1)!} \int_0^1 (1-s)^{m-1} (\nabla_X^{m-1} dJ)(X)(\exp_x(s\eta X)) ds, \quad 0 < \eta < \eta^*, \end{aligned}$$

where  $dJ : TM \rightarrow \mathbb{R}$  is a differential form of  $J$ ,  $\eta^*$  is the upper existence limit for geodesics on  $M$  and  $\nabla$  is the *Levi-Civita* connection, see [22]. Note that for compact

FIG. 6.1. Cost and state convergence on  $SE(3)$ .

manifolds  $\eta^* = \infty$ . The expansion above along the geodesic dithers in (3.4) is

$$\begin{aligned}
 J\left(\exp_x\left(\sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i}\right)\right) &= J(x) + (\nabla_{\sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i}} J)(x) + \cdots + \\
 &\frac{1}{(m-1)!} \times (\nabla_{\sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i}}^{m-1} J)(x) + \frac{1}{(m-1)!} \times \int_0^1 (1-s)^{m-1} \nabla_{\sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i}}^m \\
 &J(\exp_x s \sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i}) ds.
 \end{aligned}$$

FIG. 6.2. *State convergence on  $SE(3)$ .*

Linear properties of  $\nabla$  imply that (see [22])

$$\nabla_{\sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i}} J(x) = \sum_{i=1}^n a_i \sin(\omega_i t) \nabla_{\frac{\partial}{\partial x_i}} J(x),$$

and iteratively we have

$$\nabla_{\sum_{i=1}^n a_i \sin(\omega_i t) \frac{\partial}{\partial x_i}}^m J(x) = \sum_{i=1}^n a_i \sin(\omega_i t) \nabla_{\frac{\partial}{\partial x_i}} (\nabla_{\sum_{j=1}^n a_j \sin(\omega_j t) \frac{\partial}{\partial x_j}}^{m-1} J)(x).$$

We drop the notation  $\hat{\cdot}$  for the state trajectory in (3.5) and the dynamical equations for the extremum seeking feedback loop are given in  $x$  coordinates as follows:

$$\begin{aligned}
 \dot{x}(t) = & - \left( \sum_{i=1}^n a_i \sin(\omega_i t) J(x) \frac{\partial}{\partial x_i} + \sum_{i,j=1}^n a_i a_j \sin(\omega_i t) \sin(\omega_j t) \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i} + \cdots + \right. \\
 & \frac{1}{(m-1)!} \sum_{i,j=1}^n a_i a_j \sin(\omega_i t) \sin(\omega_j t) \times \nabla_{\frac{\partial}{\partial x_j}} \left( \nabla_{\sum_{l=1}^n a_l \sin(\omega_l t) \frac{\partial}{\partial x_l}}^{m-2} J \right)(x) \frac{\partial}{\partial x_i} + \\
 & \frac{1}{(m-1)!} \sum_{i=1}^n a_i \sin(\omega_i t) \left( \int_0^1 (1-s)^{m-1} \nabla_{\sum_{j=1}^n a_j \sin(\omega_j t) \frac{\partial}{\partial x_j}}^m \right. \\
 & \left. J \left( \exp_x \left( s \sum_{j=1}^n a_j \sin(\omega_j t) \frac{\partial}{\partial x_j} \right) \right) ds \right) \frac{\partial}{\partial x_i} \Bigg).
 \end{aligned}
 \tag{A.3}$$

Denote the new time scale by  $\tau \doteq \omega t$ , then (A.3) is a  $\tau$  varying vector field on  $(M, g^M)$  which is periodic with respect to  $\tau$ . The dynamical system (A.3) in  $\tau$  scale is given by

$$\begin{aligned}
 \frac{dx}{d\tau} = & - \frac{1}{\omega} \left( \sum_{i=1}^n a_i \sin(\bar{\omega}_i \tau) J(x) \frac{\partial}{\partial x_i} + \sum_{i,j=1}^n a_i a_j \sin(\bar{\omega}_i \tau) \sin(\bar{\omega}_j \tau) \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i} + \cdots + \right. \\
 & \frac{1}{(m-1)!} \sum_{i,j=1}^n a_i a_j \sin(\bar{\omega}_i \tau) \sin(\bar{\omega}_j \tau) \times \nabla_{\frac{\partial}{\partial x_j}} \left( \nabla_{\sum_{l=1}^n a_l \sin(\bar{\omega}_l \tau) \frac{\partial}{\partial x_l}}^{m-2} J \right)(x) \frac{\partial}{\partial x_i} + \\
 & \frac{1}{(m-1)!} \sum_{i=1}^n a_i \sin(\bar{\omega}_i \tau) \left( \int_0^1 (1-s)^{m-1} \nabla_{\sum_{j=1}^n a_j \sin(\bar{\omega}_j \tau) \frac{\partial}{\partial x_j}}^m \right. \\
 & \left. J \left( \exp_x \left( s \sum_{j=1}^n a_j \sin(\bar{\omega}_j \tau) \frac{\partial}{\partial x_j} \right) \right) ds \right) \frac{\partial}{\partial x_i} \Bigg).
 \end{aligned}$$

Let  $T$  denote the least common multiplier of the periods of  $\sin(\bar{\omega}_i \tau)$ ,  $i = 1, \dots, n$ . The average dynamical system is then given as

$$\begin{aligned}
 \frac{dx}{d\tau} = & \frac{1}{T} \int_0^T - \frac{1}{\omega} \left( \sum_{i=1}^n a_i \sin(\bar{\omega}_i \tau) J(x) \frac{\partial}{\partial x_i} + \sum_{i,j=1}^n a_i a_j \sin(\bar{\omega}_i \tau) \sin(\bar{\omega}_j \tau) \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i} + \cdots + \right. \\
 & \frac{1}{(m-1)!} \sum_{i,j=1}^n a_i a_j \sin(\bar{\omega}_i \tau) \sin(\bar{\omega}_j \tau) \times \nabla_{\frac{\partial}{\partial x_j}} \left( \nabla_{\sum_{l=1}^n a_l \sin(\bar{\omega}_l \tau) \frac{\partial}{\partial x_l}}^{m-2} J \right)(x) \frac{\partial}{\partial x_i} + \\
 & \frac{1}{(m-1)!} \sum_{i=1}^n a_i \sin(\bar{\omega}_i \tau) \left( \int_0^1 (1-s)^{m-1} \nabla_{\sum_{j=1}^n a_j \sin(\bar{\omega}_j \tau) \frac{\partial}{\partial x_j}}^m \right. \\
 & \left. J \left( \exp_x \left( s \sum_{j=1}^n a_j \sin(\bar{\omega}_j \tau) \frac{\partial}{\partial x_j} \right) \right) ds \right) \frac{\partial}{\partial x_i} \Bigg) d\tau,
 \end{aligned}
 \tag{A.4}$$

Since  $M$  is compact and  $J$  is smooth then the higher derivatives of  $J$  are all bounded above on  $M$  and (A.4) is written as

$$\frac{dx}{d\tau} = - \frac{1}{\omega} \sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_i}} J(x) \frac{\partial}{\partial x_i} + \frac{1}{\omega} \sum_{i=1}^n O((\max_{i \in \{1, \dots, n\}} a_i)^4) \frac{\partial}{\partial x_i}.
 \tag{A.5}$$

The vector field  $-\frac{1}{\omega} \sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_i}} J(x) \frac{\partial}{\partial x_i} + \frac{1}{\omega} \sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) \frac{\partial}{\partial x_i}$  is a perturbed version of the time invariant vector field  $-\frac{1}{\omega} \sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i}$  on  $(M, g^M)$ .

Following (3.7), we note that

$-\frac{1}{\omega} \sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i}$  is a scaled version of the gradient system presented in (3.7).

**Appendix B. Proof of Lemma 3.4.** Consider  $J$  as the candidate Lyapunov function on  $(M, g^M)$ . The variation of  $J$  along  $-\sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i}$  is

$$\begin{aligned} \mathcal{L}_{-\sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i}} J &= dJ \left( -\sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i} \right) \\ &= -\sum_{i=1}^n \frac{a_i^2}{2} dJ \left( \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i} \right). \end{aligned}$$

where  $\mathcal{L}$  is the Lie derivative of  $J$  along vector fields on  $(M, g^M)$ . Locally  $dJ = \sum_{i=1}^n \frac{\partial J}{\partial x_i} dx_i$ , where  $dx_i(\frac{\partial}{\partial x_j}) = \delta_{i,j}$ . Hence,

$$(B.1) \quad \mathcal{L}_{-\sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i}} J = -\sum_{i=1}^n \frac{a_i^2}{2} \left( \frac{\partial J}{\partial x_i} \right)^2 \leq 0.$$

Note that Assumption 1 guarantees that, locally,

$\sum_{i=1}^n \frac{a_i^2}{2} \left( \frac{\partial J}{\partial x_i} \right)^2 \neq 0$  for  $x \neq x^*$ , i.e.  $\mathcal{L}_f J$  is locally negative-definite.

By Assumption 1 and (B.1) the cost function  $J : M \rightarrow \mathbb{R}_{\geq 0}$  is locally positive definite, its derivative with respect to time is negative definite and  $J(x^*) = 0$ . Hence,  $J$  is a Lyapunov function on  $(M, g^M)$ , see [7]. Therefore, applying the results of [7], Theorem 6.14, implies that  $x^*$  is locally asymptotically stable.

### Appendix C. Proof of Theorem 3.5.

We analyze closeness of solutions between state trajectories of  $\frac{dx}{d\tau} = \frac{1}{\omega} f(x, \tau)$  and state trajectories of  $\frac{dx}{d\tau} = \frac{1}{\omega} \hat{f}(x)$ , where

$f(x, \tau) = -\sum_{i=1}^n \sin(\bar{\omega}_i \tau) J(\exp_{\hat{x}} \sum_{i=1}^n a \sin(\bar{\omega}_i \tau) \frac{\partial}{\partial x_i}) \frac{\partial}{\partial x_i}$  and  $\hat{f}(x) = \frac{1}{T} \int_0^T f(x, \tau) d\tau$ . We consider the periodic vector field  $Z$  defined in Lemma 3.7,  $Z(t, x) \doteq \int_0^t (\hat{f}(x) - f(x, \tau)) d\tau$ ,  $x \in M, t \in \mathbb{R}_{\geq 0}$ , where  $Z(t, x) = Z(t + T, x)$  and  $T$  is the period of the extremum seeking system  $f$ . Now consider a composition of flows on  $M$  given by

$$z(\tau) = \Phi_{\frac{1}{\omega} Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega} f}(\tau, \tau_0, x_0).$$

By the results of Lemma 3.7, the tangent vector of  $z$  is computed by

$$\begin{aligned} \dot{z}(\tau) &= T_{\Phi_{\frac{1}{\omega} f}(\tau, \tau_0, x_0)} \Phi_{\frac{1}{\omega} Z}^{(1,0)} \left( \frac{1}{\omega} f(\Phi_{\frac{1}{\omega} f}(\tau, \tau_0, x_0), \tau) \right) + \frac{\partial}{\partial \tau} (\Phi_{\frac{1}{\omega} Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega} f}(\tau, \tau_0, x_0)) \\ &= (\Phi^{-1})_{\frac{1}{\omega} Z}^{(1,0)*} \left( \frac{1}{\omega} f(\cdot, \tau) \right) (z(\tau)) + \frac{1}{\omega} \int_0^1 (\Phi^{-1})_{\frac{1}{\omega} Z}^{(1,s)*} (\hat{f}(\cdot) - f(\cdot, \tau)) ds \circ z(\tau), \end{aligned} \quad (C.1)$$

where  $(\Phi^{-1})_{\frac{1}{\omega} Z}^{(1,s)*}$  is the pullback of the state flow  $\Phi_{\frac{1}{\omega} Z}^{-1}$  and  $\epsilon = \frac{1}{\omega}$ . See Appendix D for the definition of pullbacks along diffeomorphisms. Equivalently, in a compact

form, we have

$$\begin{aligned}
 \dot{z}(\tau) &= \frac{1}{\omega} \left[ (\Phi^{-1})_{\frac{1}{\omega}Z}^{(1,0)*} f + \int_0^1 (\Phi^{-1})_{\frac{1}{\omega}Z}^{(1,s)*} (\hat{f} - f) ds \right] \circ z(\tau) \\
 (C.2) \quad &\doteq \frac{1}{\omega} H \left( \frac{1}{\omega}, \tau, z(\tau) \right).
 \end{aligned}$$

One may see that  $H(0, \tau, x) = \hat{f}(x)$  where by the construction above,  $H$  is smooth with respect to  $\frac{1}{\omega}$ . By applying the Taylor expansion with remainder we have

$$H \left( \frac{1}{\omega}, \tau, x \right) = \hat{f}(x) + \frac{1}{\omega} h(x, \zeta, \tau),$$

where  $h(x, \zeta, \tau) = \frac{\partial}{\partial \frac{1}{\omega}} H \left( \frac{1}{\omega}, \tau, x \right) |_{\frac{1}{\omega}=\zeta}$  and  $\zeta \in [0, \frac{1}{\omega}]$ . We note that  $H(\frac{1}{\omega}, \tau, x)$  is periodic with respect to  $\tau$  since  $f(x, \tau)$  and  $Z(\tau, x)$  are both T-periodic. Hence,  $h(x, \zeta, \tau)$  is a T-periodic vector field on  $M$ .

The metric triangle inequality on  $(M, g^M)$  implies

$$\begin{aligned}
 d(\Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0), \Phi_{\frac{1}{\omega}\hat{f}}(\tau, \tau_0, x_0)) &\leq d(\Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0), \Phi_{\frac{1}{\omega}Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0)) + \\
 d(\Phi_{\frac{1}{\omega}Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0), \Phi_{\frac{1}{\omega}\hat{f}}(\tau, \tau_0, x_0)) &\leq d(\Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0), \Phi_{\frac{1}{\omega}Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0)) + \\
 d(\Phi_{\frac{1}{\omega}Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0), x^*) &+ d(\Phi_{\frac{1}{\omega}\hat{f}}(\tau, \tau_0, x_0), x^*).
 \end{aligned}$$

(C.3)

Based on (C.3), We analyze the closeness of solutions for the following dynamics.

$$\begin{aligned}
 \dot{x}(\tau) &= \frac{1}{\omega} \hat{f}(x(\tau)), & x(\tau_0) &= x_0, \\
 \dot{y}(\tau) &= \frac{1}{\omega} \hat{f}(y(\tau)), & y(\tau_0) &= x_0, \\
 \dot{z}(\tau) &= \frac{1}{\omega} \hat{f}(z(\tau)) + \frac{1}{\omega^2} h(z, \zeta, \tau), & z(\tau_0) &= x_0,
 \end{aligned}$$

(C.4)

where  $\hat{f} = -\sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i}$  is the gradient system. Rescaling time back to  $t$  via  $t = \frac{1}{\omega} \tau$ , yields

$$\begin{aligned}
 \frac{dx}{dt} &= \hat{f}(x(t)), & x(t_0) &= x_0, \\
 \frac{dy}{dt} &= \hat{f}(y(t)), & y(t_0) &= x_0, \\
 \frac{dz}{dt} &= \hat{f}(z(t)) + \frac{1}{\omega} h(z, \zeta, t), & z(t_0) &= x_0,
 \end{aligned}$$

or equivalently by Lemma 3.3

$$\begin{aligned}
 \frac{dx}{dt} &= \hat{f}(x(t)) + \sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) \frac{\partial}{\partial x_i}, \\
 \frac{dy}{dt} &= \hat{f}(y(t)), \\
 \frac{dz}{dt} &= \hat{f}(z(t)) + \sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) \frac{\partial}{\partial z_i} + \frac{1}{\omega} h(z, \zeta, t),
 \end{aligned}$$

(C.5)

where  $x(t_0) = y(t_0) = z(t_0) = x_0$ .

The variation of the cost function  $J$  along  $\hat{f}(\cdot)$  is given by

$$\begin{aligned}\mathcal{L}_{\hat{f}}J &= \mathcal{L}_{\hat{f} + \sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) \frac{\partial}{\partial x_i}} J \\ &= \mathcal{L}_{\hat{f}}J + \mathcal{L}_{\sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) \frac{\partial}{\partial x_i}} J \\ &= \mathcal{L}_{\hat{f}}J + \sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) \mathcal{L}_{\frac{\partial}{\partial x_i}} J.\end{aligned}$$

As shown by the proof of Lemma 3.4, locally, we have  $\mathcal{L}_{\hat{f}}J \leq 0$ . Without loss of generality, assume positive definiteness and negative definiteness of  $J$  and

$\mathcal{L}_{\hat{f}}J = \mathcal{L}_{-\frac{1}{\omega} \sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i}} J$  are both obtained on  $U_{x^*} \subset M$  of  $x^*$ . Otherwise

we apply the intersection of the corresponding neighborhoods to perform the analysis above. Define the sublevel set  $\mathcal{N}_b$  of the positive definite function  $J : M \rightarrow \mathbb{R}_{\geq 0}$  as  $\mathcal{N}_b \doteq \{x \in M, J(x) \leq b\}$ . By  $\mathcal{N}_b(x^*)$  we denote a connected sublevel set of  $M$  containing  $x^* \in M$ . By Lemma 6.12 in [7], there exists a sublevel set  $\mathcal{N}_b(x^*) \subset U_{x^*}$ , such that  $\mathcal{N}_b(x^*)$  is compact. Consider a neighborhood of  $x^*$  denoted by  $W_{x^*}$  such that  $W_{x^*} \subset \text{int}(\mathcal{N}_b(x^*)) \subset U_{x^*} \subset M$ , where  $\text{int}()$  gives the interior set. Compactness of  $M$  implies that  $M - W_{x^*}$  is closed and compact. Hence,  $\mathcal{L}_{\hat{f}}J < 0$  for all  $x \in (M - W_{x^*}) \cap \mathcal{N}_b(x^*)$ . Note that  $\mathcal{L}_{\hat{f}}J < 0$  on  $U_{x^*} - \{x^*\}$ .

Smoothness of  $J$  and compactness of  $M - W_{x^*}$  together imply that  $\mathcal{L}_{\hat{f}}J$  attains its bounded maximum value in  $M - W_{x^*}$ . Hence, by selecting  $a_i, i = 1, \dots, n$  sufficiently small we have  $\mathcal{L}_{\hat{f}}J < 0$  on  $(M - W_{x^*}) \cap \mathcal{N}_b(x^*)$ . This implies that the state trajectory  $\Phi_{\hat{f}}(t, t_0, x_0)$  remains in  $\mathcal{N}_b(x^*)$  for  $x_0 \in \text{int}(\mathcal{N}_b(x^*))$ .

The variation of  $J$  along  $\hat{f}(z(t)) + \frac{1}{\omega}h(z, \zeta, t)$  is then given by

$$\begin{aligned}\mathcal{L}_{\hat{f} + \frac{1}{\omega}h}J &= \mathcal{L}_{\hat{f}}J + \frac{1}{\omega}\mathcal{L}_hJ \\ &= \mathcal{L}_{\hat{f}}J + \sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) \mathcal{L}_{\frac{\partial}{\partial x_i}} J + \frac{1}{\omega}\mathcal{L}_hJ.\end{aligned}$$

The same argument applies to the variation of  $J$  along  $\hat{f} + \frac{1}{\omega}h(z, \zeta, t)$  and we obtain that  $\mathcal{L}_{\hat{f} + \frac{1}{\omega}h}J < 0$  on  $(M - W_{x^*}) \cap \mathcal{N}_b(x^*)$  for sufficiently small  $a_i$  and sufficiently large  $\omega$ . Note that  $h$  is periodic with respect to  $t, \zeta \in [0, \frac{1}{\omega}]$  and  $M$  is compact. Hence,  $\mathcal{L}_hJ$  is bounded and this implies that by choosing  $a_i$  sufficiently small and  $\omega$  sufficiently large the state trajectory  $z(\cdot)$  remains in  $\mathcal{N}_b(x^*)$  for all initial states  $z_0 \in \text{int}(\mathcal{N}_b(x^*))$ .

Denote the uniform normal neighborhood of  $x^* \in M$  with respect to  $U_{x^*}$  by  $U_{x^*}^n$  (its existence is guaranteed by Lemma 5.12 in [22]). Consider a geodesic ball of radius  $\delta$  where  $U_{x^*}^n \subset \exp_{x^*}(B_\delta(0)) \subset U_{x^*}$ . By definition,  $\exp_{x^*}(B_\delta(0))$  is an open set containing  $x^*$  in the topology of  $M$ . Therefore one may shrink  $b$  to  $\hat{b}, 0 < \hat{b} \leq b$ , such that  $\mathcal{N}_{\hat{b}}(x^*) \subset \exp_{x^*}(B_\delta(0))$ . Hence, by the argument above, we select the set of initial state such as  $\Phi_{\hat{f} + \frac{1}{\omega}h}(\cdot, t_0, x_0)$  stays in a normal neighborhood of  $x^*$ . For the economy of notation we replace  $\hat{b}$  with  $b$  and assume  $\mathcal{N}_b(x^*) \subset \exp_{x^*}(B_\delta(0)) \subset U_{x^*}$ .

We analyze the distance between the state trajectory  $y(\cdot)$  of the asymptotically stable system and the averaged and full systems. As is obvious from (C.5), the averaged systems and the dynamical system corresponding to  $z(\cdot)$  are perturbations



of the gradient system  $\dot{y} = \hat{f}(y)$ , where  $\dot{y} = \hat{f}(y)$  is locally asymptotically stable and the magnitude of the perturbations vector fields is arbitrarily shrunk by adjusting  $a_i$  and  $\omega$  in (C.5). Since the initial state set is chosen such that the state trajectory  $z(\cdot)$  remains in a normal neighborhood of  $x^*$ , then the conditions of Theorems F.2 in Appendix F are satisfied. Hence, there exist a neighborhood  $U_{x^*}^1 \subset \text{int}(\mathcal{N}_b(x^*))$  and a continuous function  $\rho$ , such that for all  $x_0 \in U_{x^*}^1$

$$\limsup_{t \rightarrow \infty} d\left(\Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, x_0), x^*\right) \leq \rho\left(\sup_{z \in M, t \in [t_0, t_0+T]} \left\| \sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) \frac{\partial}{\partial z i} + \frac{1}{\omega} h(z, \zeta, t) \right\|_{g^M} \right), \quad (\text{C.6})$$

where  $\rho$  is a continuous function which is zero at zero. We note that since  $M$  is compact,  $h$  is periodic with respect to  $t$  and  $\zeta \in [0, \frac{1}{\omega}]$ , then

$\sup_{z \in M, t \in [t_0, t_0+T]} \left\| \sum_{i=1}^n O((\max_{i \in 1, \dots, n} a_i)^4) \frac{\partial}{\partial z i} + \frac{1}{\omega} h(z, \zeta, t) \right\|_{g^M}$  is bounded.

Also note that (C.6) does not guarantee the convergence of the perturbed state trajectory to  $x^*$ . However, it gives a local closeness of solutions in terms of the Riemannian distance function  $d$  to  $x^*$  after elapsing enough time. The closeness estimation provided in (C.6) is used to bound the distance between  $y(\cdot)$ ,  $x(\cdot)$  and  $z(\cdot)$  as follows. By employing the triangle inequality we have

$$\begin{aligned} d\left(\Phi_f(t, t_0, x_0), \Phi_{\hat{f}}(t, t_0, x_0)\right) &\leq d\left(\Phi_f(t, t_0, x_0), \Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, x_0)\right) + \\ (\text{C.7}) \quad &d\left(\Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, x_0), \Phi_{\hat{f}}(t, t_0, x_0)\right), \end{aligned}$$

and

$$d\left(\Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, x_0), \Phi_{\hat{f}}(t, t_0, x_0)\right) \leq d\left(\Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, x_0), x^*\right) + d\left(x^*, \Phi_{\hat{f}}(t, t_0, x_0)\right), \quad (\text{C.8})$$

where in (C.8),  $d(x^*, \Phi_{\hat{f}}(t, t_0, x_0))$  converges to zero and  $d(\Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, x_0), x^*)$  can be chosen arbitrarily small by (C.6). Note that  $\|h(z, \zeta, t)\|_{g^M}$  is bounded since  $M$  is compact,  $\zeta \in [0, \frac{1}{\omega}]$  and  $h$  is periodic with respect to  $t$ . In order to show the boundedness of  $d(\Phi_f(t, t_0, x_0), \Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, x_0))$  in (C.7), we switch back to the time scale  $\tau$ . Now we prove  $d(\Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0), \Phi_{\frac{1}{\omega}Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0)) = O(\frac{1}{\omega})$ .

By the definition of the distance function given in (2.2), we have  $d(\Phi_{\frac{1}{\omega}Z}(s, 0, x), x) \leq \ell(\Phi_{\frac{1}{\omega}Z}(s, 0, x))$ , where  $\ell(\Phi_{\frac{1}{\omega}Z}(s, 0, x))$  is the length of the curve connecting  $x$  to  $\Phi_{\frac{1}{\omega}Z}(s, 0, x)$  on  $M$ . Therefore,

$$(\text{C.9}) \quad d\left(\Phi_{\frac{1}{\omega}Z}(1, 0, x), x\right) \leq \ell(\Phi_{\frac{1}{\omega}Z}(1, 0, x)) = \frac{1}{\omega} \int_0^1 \|Z(t, \Phi_{\frac{1}{\omega}Z}(s, 0, x))\|_{g^M} ds.$$

Periodicity of  $Z$  with respect to  $t$ , boundedness of  $\Phi_{\frac{1}{\omega}Z}(\cdot, 0, x)$  in the sense of compactness of  $M$  and smoothness of  $Z$  with respect to  $x$  together yield  $d(\Phi_{\frac{1}{\omega}Z}(1, 0, x), x) = O(\frac{1}{\omega})$ . Since  $x$  is a generic element of  $M$  we have

$$d\left(\Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0), \Phi_{\frac{1}{\omega}Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0)\right) = O\left(\frac{1}{\omega}\right), \forall \tau \in [\tau_0, \infty),$$

where  $x$  is replaced by  $\Phi_{\omega f}(\tau, \tau_0, x_0) \in M$ . Hence, by using (C.3), for any  $x_0 \in U_{x^*}^1$ , there exists a time  $\hat{T}_{x_0}$ , such that

$$\begin{aligned} & d\left(\Phi_{\omega f}(\tau, \tau_0, x_0), \Phi_{\omega f}(\tau, \tau_0, x_0)\right) \leq d\left(\Phi_{\omega f}(\tau, \tau_0, x_0), \Phi_{\omega Z}^{(1,0)} \circ \Phi_{\omega f}(\tau, \tau_0, x_0)\right) + \\ & d\left(\Phi_{\omega Z}^{(1,0)} \circ \Phi_{\omega f}(\tau, \tau_0, x_0), \Phi_{\omega f}(\tau, \tau_0, x_0)\right) \leq d\left(\Phi_{\omega f}(\tau, \tau_0, x_0), \Phi_{\omega Z}^{(1,0)} \circ \Phi_{\omega f}(\tau, \tau_0, x_0)\right) + \\ & d\left(\Phi_{\omega Z}^{(1,0)} \circ \Phi_{\omega f}(\tau, \tau_0, x_0), x^*\right) + d\left(x^*, \Phi_{\omega f}(\tau, \tau_0, x_0)\right) \leq \\ & O\left(\frac{1}{\omega}\right) + \rho\left(\sup_{z \in M, t \in [t_0, t_0+T]} \left\| \sum_{i=1}^n O\left(\left(\max_{i \in 1, \dots, n} a_i\right)^4\right) \frac{\partial}{\partial z i} + \frac{1}{\omega} h(z, \zeta, t) \right\|_{g^M}\right) + \\ & d\left(x^*, \Phi_{\omega f}(\tau, \tau_0, x_0)\right), \quad \forall \tau \in [\omega \hat{T}_{x_0}, \infty). \end{aligned}$$

Note that  $\Phi_{\omega Z}^{(1,0)} \circ \Phi_{\omega f}(\tau, \tau_0, x_0) = \Phi_{f+\omega h}(t, t_0, x_0)$ , for  $\tau = \omega t$  and  $\tau_0 = \omega t_0$ . Finally we have

$$\begin{aligned} & d\left(\Phi_{\omega f}(\tau, \tau_0, x_0), x^*\right) \leq d\left(\Phi_{\omega f}(\tau, \tau_0, x_0), \Phi_{\omega f}(\tau, \tau_0, x_0)\right) + d\left(x^*, \Phi_{\omega f}(\tau, \tau_0, x_0)\right) \leq \\ & O\left(\frac{1}{\omega}\right) + \rho\left(\sup_{z \in M, t \in [t_0, t_0+T]} \left\| \sum_{i=1}^n O\left(\left(\max_{i \in 1, \dots, n} a_i\right)^4\right) \frac{\partial}{\partial z i} + \frac{1}{\omega} h(z, \zeta, t) \right\|_{g^M}\right) + \\ & 2d\left(x^*, \Phi_{\omega f}(\tau, \tau_0, x_0)\right), \quad \forall \tau \in [\omega \hat{T}_{x_0}, \infty), x_0 \in U_{x^*}^1. \end{aligned} \tag{C.10}$$

As (C.10) indicates  $d\left(\Phi_{\omega f}(\tau, \tau_0, x_0), x^*\right)$  can be ultimately bounded by shrinking  $a_i$ ,  $i = 1, \dots, n$  and increasing  $\omega$  such that the state trajectory  $\Phi_f(t, t_0, x_0)$  enters  $U_{x^*}$  and remains there.

#### Appendix D. Averaging on Riemannian manifolds.

Let us consider a perturbed system as

$$\dot{x}(t) = \epsilon f(x(t), t), \quad f \in \mathfrak{X}(M \times \mathbb{R}), x_0 \in M, \quad \epsilon \geq 0,$$

where  $f$  is periodic in  $t$  with the period  $T$ , i.e.  $f(x, t) = f(x, t+T)$ . Such a system is referred to as  $T$ -periodic. The averaged vector field  $\hat{f}$  is given by

$$\hat{f}(x) \doteq \frac{1}{T} \int_0^T f(x, s) ds, \tag{D.1}$$

where the average dynamical system is locally given by  $\dot{x}(t) = \epsilon \hat{f}(x(t))$ .

In order to obtain closeness of solutions for dynamical systems we employ the notion of pullbacks of vector fields along diffeomorphisms on  $M$  as per Definition 3.6.

We have the following lemma for the variation of smooth parameter varying vector fields.

LEMMA D.1 ([4], Page 40, [7], Page 451). *Consider a smooth parameter varying vector field  $Y(\lambda, x)$ , where  $Y \in \mathfrak{X}(\mathbb{R} \times M)$  with the associated flow  $\Phi_Y(t, t_0, \cdot) : M \rightarrow M$ . Then,*

$$\begin{aligned} & \frac{\partial}{\partial \lambda} \Phi_Y(t, t_0, x_0) = T_{x_0} \Phi_Y^{(t, t_0)} \int_{t_0}^t \left( \Phi_Y^{(s, t_0)*} \frac{\partial}{\partial \lambda} Y(\lambda, \cdot) \right) (x_0) ds = \\ & \int_{t_0}^t (\Phi_Y^{-1})^{(t, s)*} \frac{\partial}{\partial \lambda} Y(\lambda, \cdot) ds \circ \Phi_Y(t, t_0, x_0) \in T_{\Phi_Y(t, t_0, x_0)} M, \end{aligned} \tag{D.2}$$

where  $T_{x_0}\Phi_Y^{(t,t_0)}$  is the pushforward of  $\Phi_Y(t, t_0, \cdot)$  at  $x_0$ .

**D.1. Proof of Lemma 3.7.** The proof parallels the results of [7], Chapter 9 on  $\mathbb{R}^n$ . We compute the tangent vector of  $z(t) \doteq \Phi_{\epsilon Z}^{(1,0)} \circ \Phi_{\epsilon f}(t, t_0, x_0) \in M$  where  $Z(t, x) \doteq \int_0^t (\hat{f}(x) - f(x, s))ds$ ,  $0 \leq t$ . The derivative of  $\Phi_{\epsilon Z}^{(1,0)} \circ \Phi_{\epsilon f}(t, t_0, x_0)$  with respect to time has two components as follows:

$$(D.3) \quad \dot{z}(t) = T_{\Phi_{\epsilon f}(t, t_0, x_0)}\Phi_{\epsilon Z}^{(1,0)}\left(\epsilon f(\Phi_{\epsilon f}(t, t_0, x_0), t)\right) + \frac{\partial}{\partial t}(\Phi_{\epsilon Z}^{(1,0)} \circ \Phi_{\epsilon f}(t, t_0, x_0)),$$

where  $\epsilon f(\Phi_{\epsilon f}(t, t_0, x_0), t) = \frac{\partial}{\partial t}\Phi_{\epsilon f}(t, t_0, x_0)$ . The first term is the variation of  $z(t)$  with respect to the variation of the initial state  $\Phi_{\epsilon f}(t, t_0, x_0)$  and the second term is the variation of  $z(t)$  with respect to the variation of  $Z$ . Note that the flow  $\Phi_{\epsilon Z}^{(s,0)} \circ \Phi_{\epsilon f}(t, t_0, x_0)$  has two time scales  $t$  and  $s$  which are independent. Hence, the vector field  $Z$  is  $s$  invariant and  $t$  dependent where  $t$  appears as a parameter in  $Z$ .

By (3.11) we have

$$\begin{aligned} T_{\Phi_{\epsilon f}(t, t_0, x_0)}\Phi_{\epsilon Z}^{(1,0)}\left(\epsilon f(\Phi_{\epsilon f}(t, t_0, x_0), t)\right) &= (\Phi^{-1})_{\epsilon Z}^{(1,0)*}\left(\epsilon f(\cdot, t)\right)(\Phi_{\epsilon Z}^{(1,0)} \circ \Phi_{\epsilon f}(t, t_0, x_0)) \\ &= (\Phi^{-1})_{\epsilon Z}^{(1,0)*}\left(\epsilon f(\cdot, t)\right)(z(t)), \end{aligned}$$

and by Lemma D.1, we have

$$\begin{aligned} \frac{\partial}{\partial t}(\Phi_{\epsilon Z}^{(1,0)} \circ \Phi_{\epsilon f}(t, t_0, x_0)) &= \int_0^1 (\Phi^{-1})_{\epsilon Z}^{(1,s)*} \frac{\partial}{\partial t}(Z(t, \cdot))ds \circ \Phi_{\epsilon Z}^{(1,0)} \circ \Phi_{\epsilon f}(t, t_0, x_0) = \\ &= \epsilon \int_0^1 (\Phi^{-1})_{\epsilon Z}^{(1,s)*} (\hat{f}(\cdot) - f(\cdot, t))ds \circ z(t), \end{aligned}$$

Hence,

$$(D.4) \quad \dot{z}(t) = (\Phi^{-1})_{\epsilon Z}^{(1,0)*}\left(\epsilon f(\cdot, t)\right)(z(t)) + \epsilon \int_0^1 (\Phi^{-1})_{\epsilon Z}^{(1,s)*} (\hat{f}(\cdot) - f(\cdot, t))ds \circ z(t).$$

In a compact form, (D.4) is written as

$$\begin{aligned} \dot{z}(t) &= \epsilon \left[ (\Phi^{-1})_{\epsilon Z}^{(1,0)*} f + \int_0^1 (\Phi^{-1})_{\epsilon Z}^{(1,s)*} (\hat{f} - f)ds \right] \circ z(t) \\ &\doteq \epsilon H(\epsilon, t, z(t)) \in T_{z(t)}M. \end{aligned}$$

**Appendix E. Proof of Theorem 4.6.** The proof parallels the proof of Theorem 3.5 by employing the results of Lemmas 4.5 and 4.4. However, (C.9) does not necessarily hold since  $G$  is not compact. The same as (C.9) we observe that

$$\begin{aligned} d(\Phi_f(t, t_0, g_0), \Phi_{\hat{f}}(t, t_0, g_0)) &\leq d(\Phi_f(t, t_0, g_0), \Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, g_0)) + \\ &d(\Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, g_0), \Phi_{\hat{f}}(t, t_0, g_0)), \end{aligned}$$

and

$$d(\Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, g_0), \Phi_{\hat{f}}(t, t_0, g_0)) \leq d(\Phi_{\hat{f} + \frac{1}{\omega}h}(t, t_0, g_0), g^*) + d(g^*, \Phi_{\hat{f}}(t, t_0, g_0)),$$

where  $f$ ,  $\hat{f}$  and  $\hat{f}$  are the extremum seeking, averaged and gradient vector fields induced by (4.2). By the results of Lemma 4.5 and Theorem F.2,  $d(g^*, \Phi_{\hat{f}}(t, t_0, g_0))$

converges to zero and  $d(\Phi_{\hat{f}+\frac{1}{\omega}h}(t, t_0, g_0), g^*)$  is ultimately bounded by a continuous function  $\rho$  where the bound is shrunk by adjusting  $a_i, i = 1, \dots, n$  and  $\omega$ . Note that in the proof of Theorem F.2 it is shown that asymptotic stability of the gradient vector field  $\hat{f}$  guarantees that the state trajectories of  $\hat{f}$  and  $\hat{f} + \frac{1}{\omega}h$  remains in a compact subset containing the equilibrium of  $\hat{f}$ .

It remains to show  $d(\Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0), \Phi_{\frac{1}{\omega}Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0)) = O(\frac{1}{\omega})$  in the time scale  $\tau = \omega t$  for  $\tau \in [\tau_0, \infty)$ .

Following the proof of Theorem F.2 one may show  $\Phi_{\hat{f}+\epsilon h}(\omega, \omega_0, x_0) \in \mathcal{N}_b(g^*)$ ,  $\omega \in [\omega_0, \infty)$ ,  $g_0 \in \text{int}(\mathcal{N}_b(g^*))$  for some  $b \in \mathbb{R} > 0$ , where  $\mathcal{N}_b(g^*)$  is a compact connected sublevel set of a Lyapunov function containing  $g^* \in G$ , see the proof of Theorem F.2 and [7] Lemma 6.12.

Hence,  $\Phi_{\frac{1}{\omega}Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega}f}(t, t_0, x_0) \in \mathcal{N}_b(x^*)$ ,  $\tau \in [\tau_0, \infty)$ ,  $g_0 \in \text{int}(\mathcal{N}_b(g^*))$ . Therefore,

$$\begin{aligned} \bigcup_{\tau \in [\tau_0, \infty)} \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, g_0) &\subset \bigcup_{\tau \in [\tau_0, \infty)} (\Phi^{-1})_{\frac{1}{\omega}Z}^{(1,0)} \circ \mathcal{N}_b(g^*) \\ &= \bigcup_{\tau \in [\tau_0, \tau_0 + \omega T]} (\Phi^{-1})_{\frac{1}{\omega}Z}^{(1,0)} \circ \mathcal{N}_b(g^*) \\ &\subset \bigcup_{\tau \in [\tau_0, \tau_0 + \omega T], \frac{1}{\omega} \in [0, \hat{\epsilon}]} (\Phi^{-1})_{\frac{1}{\omega}Z}^{(1,0)} \circ \mathcal{N}_b(x^*), \end{aligned}$$

where the equality is due the periodicity of  $Z$  and  $\hat{\epsilon}$  is a limit for the minimum frequency ( $0 < \omega < \infty$ ). Compactness of  $\mathcal{N}_b(g^*)$ ,  $[\tau_0, \tau_0 + \omega T]$  and  $[0, \hat{\epsilon}]$  together with the smoothness of  $\Phi^{-1}$  gives compactness of  $\bigcup_{\tau \in [\tau_0, \tau_0 + \omega T], \frac{1}{\omega} \in [0, \hat{\epsilon}]} (\Phi^{-1})_{\frac{1}{\omega}Z}^{(1,0)} \circ \mathcal{N}_b(g^*)$  in  $G$ .

Hence, we show that

$$\begin{aligned} d(\Phi_{\frac{1}{\omega}Z}(1, 0, z), z) &= O\left(\frac{1}{\omega}\right), \\ \text{(E.1)} \quad z &\in \bigcup_{\tau \in [\tau_0, \tau_0 + \omega T], \frac{1}{\omega} \in [0, \hat{\epsilon}]} (\Phi^{-1})_{\frac{1}{\omega}Z}^{(1,0)} \circ \mathcal{N}_b(g^*), \end{aligned}$$

which proves

$$d(\Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0), \Phi_{\frac{1}{\omega}Z}^{(1,0)} \circ \Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0)) = O\left(\frac{1}{\omega}\right),$$

where  $z$  is replaced by  $\Phi_{\frac{1}{\omega}f}(\tau, \tau_0, x_0)$  in (E.1). The rest of the proof is identical to the proof of Theorem 3.5.

## Appendix F. Stability of perturbed systems on Riemannian manifolds.

Consider the following perturbed dynamical system on  $(M, g^M)$ .

$$\text{(F.1)} \quad \dot{x}(t) = f(x, t) + h(x, t), \quad f, h \in \mathfrak{X}(M \times \mathbb{R}).$$

The term  $h$  is considered as a perturbation of the nominal system  $f$ . The next lemma gives the existence of Lyapunov functions for dynamical systems on Riemannian manifold which satisfy specific local properties.

LEMMA F.1 ([39]). *Let  $x^*$  be an equilibrium for the smooth dynamical system  $\dot{x} = f(x, t)$  which is uniformly asymptotically stable (see [17]) on an open set  $\mathcal{N}_{x^*} \subset$*

$\mathcal{U}_{x^*}^n$  ( $\mathcal{U}_{x^*}^n$  is a normal neighborhood around  $x^*$ ). Assume  $\|T_x f(\cdot, t)\|$  is uniformly bounded with respect to  $t$  on  $\mathcal{N}_{x^*}$ , where  $\|\cdot\|$  is the norm of the bounded linear operator  $Tf : TM \rightarrow TTM$ . Then, for some  $\mathcal{U}_{x^*} \subset \mathcal{U}_{x^*}^n$ , for all  $x(t_0) = x_0 \in \mathcal{U}_{x^*}$ , there exist a differentiable function  $w : M \times \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$  and  $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathcal{K}$  (continuous, strictly increasing and zero at zero, see [17]), such that for all  $x \in \mathcal{U}_{x^*}$  and  $t \in [t_0, \infty)$ ,

$$(F.2) \quad \begin{aligned} (i) : & \alpha_1(d(x, x^*)) \leq w(x, t) \leq \alpha_2(d(x, x^*)), \\ (ii) : & \mathcal{L}_{f(x, t)} w \leq -\alpha_3(d(x, x^*)), \\ (iii) : & \|T_x w\| \leq \alpha_4(d(x, x^*)), \end{aligned}$$

where  $d(\cdot, \cdot)$  is the Riemannian metric,  $\mathcal{L}$  is the Lie derivative and  $Tw : TM \rightarrow T\mathbb{R} \simeq \mathbb{R} \times \mathbb{R}$  is the pushforward of  $w$ .

Note that the lemma above holds for the time invariant gradient system  $\dot{x} = -\sum_{i=1}^n \frac{a_i^2}{2} \nabla_{\frac{\partial}{\partial x_j}} J(x) \frac{\partial}{\partial x_i}$  since by Lemma 3.4 the gradient system is locally asymptotically stable. In this case that the Lyapunov function  $w$  is time invariant. By the results of Lemma 3.4 it has been shown that  $J$  can be considered as a Lyapunov function. However, Lyapunov  $w$  may not be necessary identical to  $J$ . One may show that items (i)-(iii) in Lemma F.1 locally hold for  $J$  around  $x^*$  when Assumption 1 is satisfied. Also note that for a compact manifold  $M$ ,  $\|T_x f\|$  is a bounded operator and the hypothesis of Lemma F.1 are satisfied for the extremum seeking algorithm (3.5) on compact manifolds.

The following theorem gives the stability of (F.1), where the nominal system is locally uniformly asymptotically stable.

**THEOREM F.2** ([39]). *Let  $x^*$  be an equilibrium of dynamical system  $\dot{x} = f(x, t)$ , which is locally uniformly asymptotically stable (see [17]) on a neighborhood  $\mathcal{N}_{x^*} \subset \mathcal{U}_{x^*}^n$  ( $\mathcal{U}_{x^*}^n$  is a normal neighborhood around  $x^*$ ). Assume the perturbed dynamical system (F.1) is complete and the Riemannian norm of the perturbation  $h \in \mathfrak{X}(M \times \mathbb{R})$  is bounded on  $\mathcal{N}_{x^*}$ , i.e.  $\|h(x, t)\|_{g^M} \leq \delta, x \in \mathcal{N}_{x^*}, t \in [t_0, \infty)$ . Then, for sufficiently small  $\delta$ , there exists a neighborhood  $\mathcal{U}_{x^*}$  and a function  $\rho \in \mathcal{K}$ , such that*

$$\limsup_{t \rightarrow \infty} d(\Phi_{f+h}(t, t_0, x_0), x^*) \leq \rho(\delta), \quad x_0 \in \mathcal{U}_{x^*}.$$

*Proof.* A full version of the proof is given in [39]. However, for the completeness of the analysis we present a sketch of the proof in this paper. Following the results of Lemma F.1, there exists  $\mathcal{U}_{x^*} \subset \mathcal{N}_{x^*}$ , such that (F.2) holds for a Lyapunov function  $w$ . By Lemma F.1, there exists  $\mathcal{U}_{x^*}$  and  $\alpha_3 \in \mathcal{K}$ , such that

$$\mathfrak{L}_{f+h} w = \mathfrak{L}_f w + \mathfrak{L}_h w \leq -\alpha_3(d(x, x^*)) + \mathfrak{L}_h w, \quad x \in \mathcal{U}_{x^*}.$$

First we show that the neighborhood  $\mathcal{U}_{x^*}$  can be shrunk, such that  $\Phi_{f+h}(t, t_0, x_0) \in \mathcal{U}_{x^*}$ ,  $t \in [t_0, \infty)$  provided  $x_0 \in \mathcal{U}_{x^*}$ . By Lemma 6.12 in [7], there exists a compact sub-level set  $\mathcal{N}_{b, t_0}(x^*) \subset \mathcal{U}_{x^*}$  where  $x^* \in \text{int}(\mathcal{N}_{b, t_0}(x^*))$ . Hence, by the Shrinking Lemma [23] there exists a precompact neighborhood  $\mathcal{W}_{x^*}$ , such that,  $\mathcal{W}_{x^*} \subset \text{int}(\mathcal{N}_{b, t_0}(x^*)) \subset \mathcal{N}_{b, t_0}(x^*)$ , see [23]. Hence,  $M - \mathcal{W}_{x^*}$  is a closed set and  $\mathcal{N}_{b, t_0}(x^*) \cap (M - \mathcal{W}_{x^*})$  is a compact set (closed subsets of compact sets are compact). The continuity of  $\alpha_3$  and  $d(\cdot, x^*)$  together with the compactness of  $\mathcal{N}_{b, t_0}(x^*) \cap (M - \mathcal{W}_{x^*})$  imply the existence of the following parameter  $\mathfrak{M}$ ,

$$\mathfrak{M} \doteq \sup_{x \in \mathcal{N}_{b, t_0}(x^*) \cap (M - \mathcal{W}_{x^*})} -\alpha_3(d(x, x^*)) < 0.$$

Note that  $\alpha_3 \in \mathcal{K}$ ,  $x \in \mathcal{N}_{b,t_0}(x^*) \cap (M - \mathcal{W}_{x^*})$  and since  $\mathcal{W}_{x^*}$  is a neighborhood of  $x^*$  then  $d(x, x^*) > 0$ ,  $x \in \mathcal{N}_{b,t_0}(x^*) \cap (M - \mathcal{W}_{x^*})$ . Therefore,  $\mathfrak{M} < 0$ . Also we have  $\mathfrak{L}_h w = dw(h) \leq \|dw\| \cdot \|h\|_{g^M} \leq \delta \|dw\|$ , where  $\|dw\|$  is the induced norm of the linear operator  $dw : TM \rightarrow \mathbb{R}$ . The smoothness of  $w$  and compactness of  $\mathcal{N}_{b,t_0}(x^*)$  together imply  $\|dw\| < \infty$ . It is important to note that  $\|dw\|$  is closely related to  $\|Tw\|$  through the component of the Riemannian metric  $g$ . As is shown by Theorem F.1,  $\|T_x w\| \leq \alpha_4(d(x, x^*))$ . Hence, the smoothness of  $M$  and compactness of  $\mathcal{N}_{b,t_0}(x^*)$  imply that  $\|dw\| < \infty$ . Note that  $\|dw\|$  is the norm of the linear operator  $dw : T_x M \rightarrow \mathbb{R}$ . Hence, for sufficiently small  $\delta$ , we have  $\mathfrak{L}_{f+h} w < 0$ ,  $x \in \mathcal{N}_{b,t_0}(x^*) \cap (M - \mathcal{W}_{x^*})$ . Therefore, the state trajectory  $\Phi_{f+h}(t, t_0, x_0)$ ,  $t \in [t_0, \infty)$  stays in  $\mathcal{U}_{x^*}$  for all  $x_0 \in \text{int}(\mathcal{N}_{b,t_0}(x^*))$ .

Without loss of generality, we assume  $\mathcal{U}_{x^*} = \exp_{x^*} B_{r_1}(0)$ ,  $r_1 < i(x^*)$ . Note that for sufficiently small  $r_1$ , we have  $\exp_{x^*} B_{r_1}(0) \subset \mathcal{N}_{b,t_0}(x^*)$ . Then, by the results of Lemma F.1, the variation of  $w$  along  $f + h$  is given by

$$\begin{aligned} \mathfrak{L}_{f+h} w &= \mathfrak{L}_f w + \mathfrak{L}_h w \leq -\alpha_3(d(x, x^*)) + \mathfrak{L}_h w \\ &= -\alpha_3(d(x, x^*)) + dw(h(x, t)) \\ &= -\alpha_3(d(x, x^*)) + T_x w(h(x, t)) \\ &\leq -\alpha_3(d(x, x^*)) + \|T_x w\| \cdot \|h(x, t)\|_{g^M} \\ &\leq -\alpha_3(d(x, x^*)) + \delta \alpha_4(d(x, x^*)) \\ &\leq -(1 - \theta) \alpha_3(d(x, x^*)) - \theta \alpha_3(d(x, x^*)) \\ &\quad + \delta \alpha_4(d(x, x^*)) \leq -(1 - \theta) \alpha_3(d(x, x^*)), \\ &\text{if } \alpha_3^{-1}\left(\frac{\delta \alpha_4(r_1)}{\theta}\right) \leq d(x, x^*) \leq r_1, \end{aligned}$$

for some  $r_1 < i(x)$ ,  $0 < \theta < 1$ .

Define  $\eta \doteq \alpha_2(\alpha_3^{-1}(\frac{\delta \alpha_4(r_1)}{\theta}))$ , then  $\{x \in M \mid d(x, x^*) \leq \alpha_3^{-1}(\frac{\delta \alpha_4(r_1)}{\theta})\} \subset \mathcal{N}_{t,\eta} = \{x \in \mathcal{U}_{x^*} \mid w(x, t) \leq \eta\} \subset \{x \in M \mid \alpha_1(d(x, x^*)) \leq \eta\}$ . Hence, solutions initialized in  $\{x \in M \mid d(x, x^*) \leq \alpha_3^{-1}(\frac{\delta \alpha_4(r_1)}{\theta})\}$  remain in  $\{x \in M \mid \alpha_1(d(x, x^*)) \leq \eta\}$  since  $\dot{w} < 0$  for  $x \in \mathcal{N}_{t,\eta} - \{x \in M \mid d(x, x^*) \leq \alpha_3^{-1}(\frac{\delta \alpha_4(r_1)}{\theta})\}$ . This proves

$$\begin{aligned} \limsup_{t \rightarrow \infty} d(\Phi_{f+h}(t, t_0, x_0), x^*) &\leq \alpha_1^{-1} \left( \alpha_3^{-1} \left( \frac{\delta \alpha_4(r_1)}{\theta} \right) \right) \\ &\doteq \rho(\delta), \end{aligned}$$

for any  $x_0 \in U_{x^*} \doteq \{x \in M \mid d(x, x^*) < \alpha_3^{-1}(\frac{\delta \alpha_4(r_1)}{\theta})\} \cap \text{int}(\mathcal{N}_{b,t_0}(\hat{x}))$ .  $\square$

## REFERENCES

- [1] R. Abraham, J. E. Marsden, and T. S. Ratiu. *Manifolds, Tensor Analysis, and Applications*. Springer, 1988.
- [2] P.A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2007.
- [3] M. Adachi. *Embeddings and Immersions*. American Mathematical Soc., 1993.
- [4] A. Agrachev and Y. Sachkov. *Control Theory from the Geometric Viewpoint*. Springer, 2004.
- [5] C. G. Baker, P.-A. Absil, and K. A. Gallivan. An implicit trust-region method on Riemannian manifolds. *IMA J. Numer. Anal.*, 28:665–689, 2008.
- [6] A. M. Bloch. *Nonholonomic Mechanics and Control*. Springer, 2000.
- [7] F. Bullo and A.D. Lewis. *Geometric Control of Mechanical Systems: Modeling, Analysis, and Design for Mechanical Control Systems*. Springer, 2005.
- [8] H. B. Dürr, M. S. Stanković, C. Ebenbauer, and K. H. Johansson. Lie bracket approximation of extremum seeking systems. *Automatica*, (49):1538–1552, 2013.

- [9] H. B. Dürr, M. S. Stanković, K. H. Johansson, and C. Ebenbauer. Extremum seeking on submanifolds in Euclidean spaces. *Automatica*, Provisionally accepted for Publication, 2014.
- [10] O. Freifeld. Statistics on manifolds with applications to modeling shape deformations. *PhD thesis, The Division of Applied Mathematics at Brown University*, 2014.
- [11] P. Frihauf, M. Krstić, and T. Bašar. Nash equilibrium seeking in non cooperative games. *IEEE Trans. Automatic Control*, 57(5):1192–1207, 2012.
- [12] D. Gabay. Minimizing a differentiable function over a differentiable manifold. *Journal of Optimization Theory and Applications*, 37:177–219, 1982.
- [13] M. Guay, D. Dochain, and M. Perrier. Adaptive extremum seeking control of nonlinear dynamic systems with parametric uncertainties. *Automatica*, 39:1283–1293, 2003.
- [14] M. Guay, D. Dochain, and M. Perrier. Adaptive extremum seeking control of continuous stirred tank bioreactors with unknown growth kinetics. *Automatica*, 40:881–888, 2004.
- [15] U. Helmke, S. Riardo, and J. B. Yoshizawa. Newtons algorithm in Euclidean jordan algebras, with applications to robotics. *Communications in Information and Systems*, 3(2):283297, 2002.
- [16] J. Jost. *Riemannian Geometry and Geometrical Analysis*. Springer, 2004.
- [17] H. K. Khalil. *Nonlinear Systems*. Prentice Hall, 2002.
- [18] S. Z. Khong, D. Nešić, C. Manzie, and Y. Tan. Multidimensional global extremum seeking via the direct optimisation algorithm. *Automatica*, 49(7):1970–1978, 2013.
- [19] W. P. A. Klingenberg. *Riemannian Geometry*. de Gruyter Studies in Mathematics, 1995.
- [20] A. W. Knap. *Lie groups Beyond an Introduction*. Birkhauser, 1996.
- [21] M. Krstić and H. W. Wang. Stability of extremum seeking feedback for general nonlinear dynamic systems. *Automatica*, 36:595–601, 2000.
- [22] J. M. Lee. *Riemannian Manifolds, An Introduction to Curvature*. Springer, 1997.
- [23] J. M. Lee. *Introduction to Topological Manifolds*. Springer, 2000.
- [24] J. M. Lee. *Introduction to Smooth Manifolds*. Springer, 2002.
- [25] D. G. Luenberger. The gradient projection method along geodesics. *Management Science*, 18(11):620–631, 1972.
- [26] J. H. Manton. Optimisation algorithms exploiting unitary constraints. *IEEE Transactions on Signal Processing*, 50(3):635650, 2002.
- [27] D. Nešić, A. Mohammadi, and C. Manzie. A framework for extremum seeking control of systems with parameter uncertainties. *IEEE Trans. Aut. Cont.*, 58(2):435–448, 2013.
- [28] D. Nešić, A. Mohammadi, and C. Manzie. A systematic approach to extremum seeking based on parameter estimation. In *IEEE Conf. Decision and Control*, pages 3902–3907, Atlanta, 2010.
- [29] D. Nešić, Y. Tan, W. Moase, and C. Manzie. A unifying approach to extremum seeking: adaptive schemes based on estimation of derivatives. In *IEEE Conf. Decision and Control*, pages 4625–4630, Atlanta, 2010.
- [30] X. Pennec. Bi-invariant means on Lie groups with Cartan-Schouten connections. *Lecture Notes in Computer Science, Geometric Science of Information*, 8085:59–67, 2013.
- [31] P. Petersen. *Riemannian Geometry*. Springer, 1998.
- [32] W. Ring and B. Wirth. Optimization methods on Riemannian manifolds and their application to shape space. *SIAM journal on Control and Optimization*, 22(2):596–627, 2012.
- [33] W. Rudin. *Real and Complex Analysis*. New York: McGraw-Hill, 1974.
- [34] J. A. Sanders and F. Verhulst. *Averaging Methods in Nonlinear Dynamical Systems*. Springer, 1985.
- [35] B. O. Shubert. A sequential method seeking the global maximum of a function. *SIAM Journal on Numerical Analysis*, 17(9):379388, 1972.
- [36] S. T. Smith. Optimization techniques on Riemannian manifolds. *Fields Institute Communications*, 3(3):113–135, 1994.
- [37] R. G. Strongin and Y. D. Sergeyev. *Global Optimization with Non-Convex Constraints*. Springer Science+Business Media Dordrecht, 2000.
- [38] Y. Tan, D. Nešić, and I. M. Y. Mareels. On non-local stability properties of extremum seeking control. *Automatica*, 42(6):889–903, 2006.
- [39] F. Taringoo, P. M. Dower, D. Nešić, and Y. Tan. *A Local Characterization of Lyapunov Functions and Robust Stability of Perturbed Systems on Riemannian Manifolds*, <http://arxiv.org/abs/1311.0078>. Submitted to *Automatica*, 2013.
- [40] A. R. Teel and D. Popović. Solving smooth and non-smooth multivariable extremum seeking problems by the methods of nonlinear programming. In *American Control Conference*, pages 2394–2399, 2001.
- [41] C. Udriste. *Convex Functions and Optimization Methods on Riemannian Manifolds*. Kluwer

- Academic Publishers, 1994.
- [42] V. Varadarajan. *Lie groups, Lie algebras, and their representations*. Springer, 1984.
  - [43] Y. Yang. Globally convergent optimization algorithms on Riemannian manifolds: Uniform framework for unconstrained and constrained optimization. *Journal of Optimization Theory and Applications*, 132(2):245–265, 2007.